MĀORI AS PHRASE-BASED LANGUAGE

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI‘I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

LINGUISTICS

DECEMBER 2014

By

Fumiko S. Yamada

Dissertation Committee:

Yuko Otsuka, Chairperson
   Rapata Wiri
   Robert Blust
   William O’Grady
   Gregory Maskarinec

Keywords: Māori, parts of speech, categories
Table of Contents

List of Abbreviations .................................................................................................................. v

Chapter 1: Introduction
1.1. Introduction .......................................................................................................................... 1
1.2. Research Questions .............................................................................................................. 2
1.3. Hypothesis and Goal ........................................................................................................... 5
1.4. Methodology ....................................................................................................................... 7
1.5. Contributions and Implications .......................................................................................... 8

Chapter 2: History of Parts of Speech
2.1. Introduction .......................................................................................................................... 11
2.2. The Origin of the Part-of-Speech System ........................................................................... 11
2.3. Word Classification for the American Structuralists .......................................................... 14
2.4. Categories as Universal ....................................................................................................... 17
2.5. Confusion Regarding Categories ....................................................................................... 19
2.6. Categories as Features ........................................................................................................ 22
2.7. Categories as Prototypes .................................................................................................... 27
    2.7.1. Against syntax-based word classification ................................................................. 27
    2.7.2. Language as a part of cognition ................................................................................. 28
2.8. Parts of Speech in Typology ............................................................................................... 34
2.9. Conclusion .......................................................................................................................... 40

Chapter 3: Issues in Māori Grammar
3.1. Introduction .......................................................................................................................... 43
3.2. Basic Structure of Māori Words and Phrases ................................................................. 43
3.3. Basic Morphosyntax of Māori .......................................................................................... 43
    3.3.1. Determiners .................................................................................................................. 46
    3.3.2. Prepositional particles .............................................................................................. 49
    3.3.3. TAM ............................................................................................................................ 52
3.4. Māori Parts of Speech

3.4.1. Lexical items in nominal phrases

3.4.2. Lexical items in verbal phrases

3.4.3. Parts of speech or not?

3.5. -Canga and -Cia

3.5.1. Nominalizing suffix -Canga

3.5.2. Passive suffix -Cia

3.5.3. Interim summary

3.6. Proposed Word Classification Systems for Māori

3.6.1. Biggs’s analysis

3.6.2. Bauer’s analysis

3.6.3. Harlow’s analysis

3.7. Parts of Speech or Not?

Chapter 4: Category Undespecification

4.1. Introduction

4.2. Morphology without Categories

4.3. Foley (Philippine Languages)

4.4. Bisang (Late Archaic Chinese)

4.5. Broschart (Tongan)

4.6. Peterson (Kharia)

4.7. Gil (Riau Indonesian)

4.8. Biggs (Māori)

4.9. Conclusion

Chapter 5: Māori as Phrase-based Language

5.1. Māori Phrases as Basic Grammatical Units

5.2. The Classic Part-of-Speech System and Māori Grammar

5.3. Prototype Categories in Māori

5.4. Consequences of Identifying Parts of Speech
5.5. Facts of Māori ........................................................................................................... 118
5.6. Base Category as a phrase ....................................................................................... 119
5.7. Category-less Bases ................................................................................................. 124
5.8. Māori as a Type/Token Language ........................................................................... 127
5.9. Māori Phrase Structure Revisited .......................................................................... 141
5.10. Affixes as Type ........................................................................................................ 147
  5.10.1. -Canga as [-pred] ............................................................................................... 150
  5.10.2. -Cia as [+high transitivity][+perfective][+affectedness of DO] ............................... 152
  5.10.3. Double suffixation .............................................................................................. 155
5.11. Conclusion ............................................................................................................... 158

Chapter 6: Conclusion ..................................................................................................... 160

Bibliography .................................................................................................................... 171
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABS</td>
<td>absolutive</td>
<td>PST</td>
<td>past</td>
</tr>
<tr>
<td>ANAPH</td>
<td>anaphoric</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ART</td>
<td>article</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT</td>
<td>dative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEF</td>
<td>definitive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DET</td>
<td>determiner</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DO</td>
<td>direct object</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXCL</td>
<td>exclusive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FUT</td>
<td>future</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GEN</td>
<td>genitive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INCL</td>
<td>inclusive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NMLZ</td>
<td>nominalize</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OBL</td>
<td>oblique</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>preposition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC</td>
<td>locative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PASS</td>
<td>passive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PERS</td>
<td>personal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POSS</td>
<td>possessive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRES</td>
<td>present</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRST</td>
<td>presentative/essive</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1. Introduction

Māori is an indigenous language spoken by the Māori people of Aotearoa (New Zealand). It belongs to the Polynesian subgroup of Oceanic languages in the Austronesian language family. The 2013 census conducted by the New Zealand government shows that approximately twenty percent of the country’s citizens can hold a basic conversation in te reo Māori (Māori language), which is about a five percent decrease from the 2006, and one in five Māori speaks more than one language (Statistics New Zealand Tatauranga Aotearoa 2013).

Māori shares many linguistic characteristics with other Polynesian languages such as Samoan and Tongan, including a simple phoneme inventory (containing ten consonants and five vowels) and predicate-initial sentence structure. Another such characteristic is that the same lexical item may be used as an adjective, noun, or verb, as shown in (1.1–1.3). So-called “parts of speech” are fluid in Māori.

(1.1) He tangata mōmona ahau.
DET man fat 1SG¹
‘I am a fat man.’

(1.2) He mōmona ahau.
DET fat 1SG
‘I am fat.’

¹ A list of glosses used in this dissertation appears in List of Abbreviatioon. The glosses of examples from other authors have been changed to conform with the glossing system used in this dissertation.
(1.3) Kei te mōmona ahau.
     TAM fat 1SG
     'I am fat' or 'I am in a state of being fat.'

In (1.1), mōmona is used as a postnominal modifier, denoting property, that is, an adjective. The same word mōmona is preceded by a determiner he in (1.2), suggesting that it is used as a noun. Mōmona can also occur after a tense, aspect, and mood (TAM) marker such as kei te in (1.3). This suggests that mōmona is used as a verb in (1.3), for tense, aspect, and mood are usually properties of verbs. In these three different distributions, the word mōmona can be used without changing its form.

1.2. Research Questions

The discussion of the word mōmona in section 1.1 includes a brief observation about each of the three usages shown in the examples: In (1.1), mōmona is used as a modifier, denoting property; in (1.2), it is analyzed as a noun, because it is preceded by a determiner; and in (1.3), it follows TAM, and thus it is a verb. These observations are acceptable, or at least comprehensible, because the writer and the reader share a common knowledge of the terminology ("noun," "verb," "adjective," "modifier," and so forth), and we know more or less what the terms mean. Although this situation may not seem unusual, it is a matter of great interest to me that nobody seems to see anything wrong with using category-denoting terms ("noun/verb/adjective") to analyze Māori data. That this is so demonstrates how completely these categories are accepted as universal tools for grammatical description, applicable to any language. Because the part-of-speech system is so widely accepted as a universally applicable word classification method, the result of an analysis of Māori data shows that the same word,
mōmona ‘fat’, can be used as a noun, verb, and adjective: Māori word classes overlap. The phenomenon of overlapping categories has been observed in other languages: Tongan (Broschart 1997), Fijian (Schütz 1975), Samoan (Hengeveld, Rijkoff, and Siewierska 2005), Riau Indonesian (Gil 2005), Kharia (Peterson 2013), and Mundari (Evans and Osada 2005), to name a few. Two possible scenarios immediately suggest themselves to explain the phenomenon in (1.1–1.3):

1. Māori does not possess a classic noun/verb/adjective part-of-speech distinction. Instead, it has only a single word class that contains all nouns, verbs, and adjectives.

2. Māori does have a noun/verb/adjective distinction, but the word forms do not change; that is, the same form can appear freely in nominal, verbal, and adjectival contexts.

Neither of these proposals questions the existence of the part-of-speech system as a universal part of grammatical description. If, however, we approach the task of accounting for Māori grammar without taking the necessity of the parts of speech for granted, we may well come to doubt the usefulness of the part-of-speech system as a word classification method for Māori.

The idea of parts of speech first emerged in the third or fourth century BCE, articulated by Greek philosophers and philologists. Plato recognized that the sentence is a combination of two different elements, ónoma (‘subject’) and rhême (‘predicate’); Aristotle distinguished verbs and nouns. In the first century BCE, a Greek grammar differentiated the eight parts of speech, primarily based on the inflectional system with notional definitions. The part-of-speech system developed by the Greek scholars for the Greek language was inherited by Latin grammarians and later by scholars of other European languages. Not only is this European tradition of word classification still in use, but its validity is firmly believed.

It is not surprising that a Greek-origin, morphology-based word classification system has
limitations when it is applied to Māori, which has no inflectional morphemes. Because the word forms remain the same in Māori in nominal, verbal, and adjectival uses, other criteria have to be used for word class identification. Therefore, this investigation into the overlapping Māori word categories asks the following descriptive questions:

1. Is the classic noun/verb/adjective distinction applicable to Māori?

2. If it is found to be applicable, can the conventional criteria accurately capture the facts of Māori?

3. If it is not applicable, what problems does Māori raise for the traditional classification of parts of speech?

4. What are alternative solutions for Māori?

Meanwhile, the fact that the part-of-speech system has survived over two thousand years does suggest the universality of the parts of speech, or of the classes of noun, verb, and adjective. Therefore, the following theoretical questions also need to be asked:

1. What is the role of parts of speech (or word categories) in modern linguistics?

2. Are the parts of speech (i.e., nouns, verbs, adjectives) universal or language-specific?

3. Are parts of speech (or word categories) an inherent property of the word?

These questions in turn raise new questions. For instance, in order to seek the answers, one must clearly define the parts of speech. This necessity leads to related questions such as, “Are the parts of speech pre-theoretical, indefinable primitives? How are they assigned during the course of word derivation? Do the lexical items in the lexicon all have the categories?” and so on.

In the end, the question that this dissertation asks is what significance the Māori word class overlap phenomenon has for the descriptive and theoretical fields in the study of languages. The immediate implication of the category overlapping is that the part-of-speech system has
limitations and that word classification is language-specific.

1.3. Hypothesis and Goal

Data such as examples (1.1–1.3) that show the overlapping nature of the Māori word categories raise the question of what it means that the categories overlap. If the words are categorized for grammatical reasons, such categories should not overlap. If the language is organized in such a way that the words do not need categories, then this language has no word categories to overlap in the first place. The central hypothesis of this dissertation is that word categorization occurs at the grammatical level that is necessary for the language. Hence, word categories are language-specific. Instead of asking if Māori has parts of speech, a different approach is required to accurately describe its word classification. Broschart (1997:160) argues as follows:

What is needed, in any case, are adequate, in-depth studies which force us to look at the distinctions languages really make, rather than to look for distinctions only because we have inherited them from the study of Indo-European.

As explained earlier, the part-of-speech system, or the noun/verb/adjective distinction, is the product of scholarly work based on Classical Greek, which has an elaborate inflectional system for nouns as well as verbs. Therefore, the inflectional morphemes serve as category indicators. The morphological criterion, that is, the use of inflectional morphemes as indicators of parts of speech, has limitations, because it obviously cannot be applied to languages such as Māori that have few or no inflectional morphemes. Broschart’s argument suggests that Māori word classification should be the product of the grammatical description of Māori, rather than
the grammatical description of Māori being a product of the classical part-of-speech distinction.

The goal of this investigation is to show that different word categories exist in different languages, and for Māori, the objects of classification are not individual lexical items such as mōmona, but rather entire phrases such as he tangata mōmona, he mōmona, and kei te mōmona in (1.1–1.3). An analysis that rejects the categories on the lexical level has been proposed for Tongan by Broschart (1997), and for Kharia (a Munda language spoken in India and Bangladesh) by Peterson (2013), for instance. This analysis is also consistent with one statement that is agreed on among Māori linguists: The phrase is the basic grammatical unit in Māori. It follows that the category that is relevant to Māori grammar is not the lexical category, but the syntactic (phrasal) category.

The stance taken in this dissertation is inspired by Peterson (2013), who claims in his study of Kharia that the category overlap phenomenon can be accounted for by using a grammatical model with the noun/verb/adjective distinction or by using a category underspecification model. Peterson concludes that what makes one account preferable to the other is greater simplicity, based on the principle of Occam’s Razor. Māori grammar can be described in a framework that posits distinct nouns, verbs, and adjectives as universal categories.

The point of view taken in this dissertation is that such a distinction is not necessary to account for Māori grammar.

This research is also motivated by Lehmann (2013:59), who writes:
Languages are problem-solving systems. Some of the tasks to be solved are universal. The task of assigning expressions to categories is among these, so it must be incorporated into linguistic theory. However, incorporating one of the possible solutions into linguistic theory fails to recognize that the solution of a
problem is dynamic in nature and there are generally alternative ways of solving a problem. The task of universals research is not to stipulate one of the possible solutions as a property of universal grammar, but to analyze the variation encountered in order to identify its principle.

This interpretation brings our attention back to the starting point in the study of parts of speech, or word categories, which is the role of categorization in grammar. All languages have categories that serve their purpose, which is to form sentences successfully. Māori has its own solution to the problem of categorization, and the Māori categories exist to fulfill their purpose: deriving well-formed Māori sentences.

1.4. Methodology

To answer the research questions discussed in section 1.2, I use the following methodologies. First, Māori data were collected from three sources:

1. Texts including reference grammars, dictionaries, a corpus collected by Dr. Mary Boyce, and scholarly articles. These sources appear in the reference list.

2. Email communication with Dr. Rapata Wiri at Waiaariki Institute of Technology

3. Consultation with Dr. Mary Boyce at the University of Hawai‘i at Mānoa

Second, the parts of speech are explored by investigating their origin and development in the history of the study of languages. Major theoretical models such as American structuralism, generative grammar, cognitive grammar, and construction grammar are discussed to compare how they treat the issue of the parts of speech. Some of my theoretical questions will be answered by researching both the theories that assume that lexical categories are an inherent part
of lexical items and those that do not assume that lexical categories are linguistic primitives given to every lexical item.

In order to find possible alternative solutions to Māori word categories, previous research that deals with similar phenomena will be examined. Other Austronesian languages such as Tongan, Niuean, and Philippine languages are known to have flexible word categories. Kharia does not show a distinction between noun and verb. Late Archaic Chinese lost the morpheme that showed category. Riau Indonesian is reported to have only one syntactic category. These languages are best described not only without positing the classic part-of-system, but also by differentiating the lexical and syntactic levels for category assignment. Although it has been believed that each word receives a category (Schacter 1985), it is necessary to re-examine the basic nature of the word categories.

Finally, Māori data will be examined by using a theoretical framework that assumes the existence of lexical categories, using notations such as N(oun), V(erb), and A(djective), and it also will be examined using an approach that assumes lexical items to be category-less.

1.5. Contributions and Implications

The question of how to identify and define parts of speech in Māori is twofold. First, there is a problem of methodology. The traditional part-of-speech distinction, based on Classical Greek, clearly has limitations when applied to Māori, a language without Greek-style inflection. This has led linguists to question whether the traditional part-of-speech distinction exists in Māori. The real issue, however, is not whether Māori distinguishes between nouns and verbs. Rather, it is how Māori words are categorized. Therefore, it is essential to assess the appropriateness of criteria used to define parts of speech and establish a set of criteria suitable for Māori.
Second, there is the question of whether the objects of investigation are the right kind of object. The part-of-speech system was created based on the assumption that a “word” is a grammatical unit. “Word” is an elusive term to define, however. If words are taken to be the building blocks of sentences, the first task is to determine whether this is also the case in Māori. It has been proposed that the basic grammatical unit in Māori is not the individual lexical item, but the phrase (Bauer 1993, 1997; Biggs 1961, 1969; Harlow 2001, 2007). If this is correct, it is meaningless to persist in discussing whether an individual lexical item such as mōmona ‘fat’ by itself is a verb or noun or adjective. Rather, the relevant task is to identify and define the parts of speech (or the categories) at the phrasal level. As the problem of fluidity and/or non-existence of parts of speech has come to be recognized in many other languages, a careful examination of Māori word categorization has cross-linguistic implications.

The hypothesis that the categories are assigned to phrases rather than lexical items in Māori also has theoretical implications. In some theoretical frameworks such as generative grammar, syntactic categories are synonymous with lexical categories. However, Māori data suggest that syntactic and lexical categories represent two distinct notions for languages such as Māori. Due to the lack of inflectional morphemes, word forms cannot be used as indicators of the categories. Therefore, the determining factor in categorizing Māori words is their distribution, which raises an important issue. Namely, a part of speech, or a lexical category, cannot be an inherent part of a lexical item, at least in some languages.

1.6. Organization

This dissertation is organized in the following manner. In chapter 2, the history of the parts of speech will be discussed. American structuralism, generative grammar, and cognitive
and construction grammar will be introduced to illustrate their differences in treating the parts of speech or word categories. Cognitive grammar, in particular, offers an alternative interpretation of the categories. Typological approaches, which explore the universal implications of the types of part-of-speech systems, will also be discussed. Chapter 3 will describe and illustrate the category overlap phenomenon in Māori, and present the approaches taken to the issue by the major Māori grammarians. Chapter 4 will introduce several analyses of languages with flexible word classes that do not posit lexical categories. Chapter 5 will present Māori data, considering it first with an analysis based on the noun/verb/adjective distinction, and then with the alternative analysis. I will conclude in chapter 6 that the categories relevant to Māori grammar are the most appropriate categories, and I will discuss the advantage of taking an approach without the noun/verb/adjective distinction.
Chapter 2

History of Parts of Speech

2.1. Introduction

In this chapter, I will describe how the concept of parts of speech was introduced into the study of languages and how it has come to be interpreted in various ways. In 2.2, I will briefly discuss the origin of the parts of speech in classical Greek thought. In 2.3, I will describe how the American structuralists challenged the idea of parts of speech and offered a new and alternative view of languages and the purpose of the study of languages. In 2.4, 2.5, and 2.6, I will highlight the changes that generative grammar brought to the interpretation of parts of speech. I will discuss some post-generative approaches to parts of speech in 2.7, and in 2.8 I will examine the parts of speech in terms of language typology. The chapter concludes in 2.9.

2.2. The Origin of the Part-of-Speech System

According to Campbell (2001), Luhtala (2005), and Rauh (2010), the idea of parts of speech emerged first in the works of Greek philologists and philosophers before the first century BC, and was then inherited and developed further by Latin grammarians. The philologists who began to study the Greek language were motivated by the desire to preserve writings from the past, while the philosophers, such as Plato and Aristotle, aimed at the analysis of language as the expression of thought. Plato is the first scholar who recognized ónoma, “the articulate sign set on those who do the actions” and rhême, “that which denotes action” (Sandys 1903:90), which are equivalent to “subject” and “predicate,” respectively. Plato also explained that these two elements have to be “mingled together” for language to be formed. Therefore, Plato was aware
of the syntactic aspect of language. Aristotle further distinguished verbs, as connoting time, from nouns, which do not. The eight parts of speech\(^2\) are mentioned already in a Greek grammar written in the first century BC, *Tēkhnē grammatikē* by Dionysius Thrax. This grammar’s primary criteria for word classification are morphological, but it also considers semantic (notional) criteria. For instance, a noun is “a part of the sentence which is subject to case inflection, and signifies something corporeal or non-corporeal; by corporeal I mean something like ‘a stone’, and by non-corporeal something like ‘education’” (Thrax 1986:350), and a verb is a “word which is without case inflection, displaying changes of tense, person, and number, and signifying also activity or passivity” (Thrax 1986:354) Adjectives are not an independent part of speech but fall under the category of nouns according to this morphologically based classification, because they have the same inflections as nouns. Although Plato had already introduced the notion of syntax—that is, that logically well-formed sentences must have two essential parts—the popularity of the *Tēkhnē grammatikē* made the morphology- and semantic-oriented part-of-speech system the dominant theory in the study of language for later generations. The *Tēkhnē grammatikē* was translated into Latin, which influenced the description of Latin grammar. Grammarians such as Aleius Donatus (mid-fourth century), who wrote *Ars Grammatica*, and Priscian Caesariensis (early sixth century) who wrote *Institutiones Grammaticae* (Robins 1976:19), maintained the eight parts of speech, which were defined notionally, although with some reference to inflection, and using the terminology of the *Tēkhnē grammatikē*. The adaptation of Greek grammar to Latin shaped the descriptions of the Romance languages produced by later generations of grammarians, who carried on using part-of-speech systems.

\(^2\) They are: noun, verb, participle, article, pronoun, preposition, adverb, and conjunction.
As mentioned in the previous section, the Greek philosophers were motivated to study language because they viewed language as the expression of thought, while the philologists were more concerned with the documentation and the preservation of literary works. The philosophers’ viewpoint was shared by Antoine Arnauld and Claude Lancelot, who wrote the *Port-Royal Grammar* for the abbots of the monastery of Port-Royal in France in the seventeenth century (Campbell 2001: 85). The *Port-Royal Grammar* is significant to modern linguistics, according to Lakoff (1975:348), because it is the “spiritual ancestor” of Chomky’s (1957) *Syntactic Structures*. The *Port-Royal Grammar* is heavily influenced by the Cartesian philosophy of human intellectual faculties and logic. In the Cartesian way of thinking, ideas exist before words, and languages are tools that reflect human minds. This view of languages gave more importance to what words signify, that is, semantics, than did the Greek grammar in which word forms were the primary concern. The *Port-Royal Grammar* did, however, carry on the Greek tradition of parts of speech in spite of differences in the languages and the justifications for classifying words. The parts of speech, according to the Port-Royal grammarians, belonged to two basic categories: objects of thoughts and manners of thoughts. Nouns are words that express objects of thoughts, while verbs express the manner of thoughts. Although word form is not an important criterion in the *Port-Royal Grammar*, the traditional morphology-based part-of-speech system was at least not a hindrance to its analysis of French, and following the Greek and Latin tradition, it did not use syntactic criteria for word classification. The motivation for the *Port-Royal Grammar*, originally titled *Grammaire Générale et Raisonné*, was to unravel the relationship between language and mind. While the Greek philologists aimed at documenting the literary work of past scholars, the Port-Royal grammarians sought a “general and rational” grammar that would represent the human mental process and therefore be universal. These two
different views of language were passed on in the field of linguistics in later generations, and the part-of-speech system, created by Greek scholars for the Greek language and later adopted for the analysis of Romance languages, was also passed on, enduring in the study of languages to the present day.

2.3. Word Classification for the American Structuralists

The American structuralist movement started in the field of anthropology, which was involved in the investigation of Native American cultures. The anthropologist Franz Boas realized that Native American languages (i.e., non-Indo European languages) could not be adequately described by traditional grammar, which was created and developed for Indo-European languages. Regarding grammatical categories, Boas (1911:35) wrote:

Grammarians who have studied the languages of Europe and western Asia have developed a system of categories which we are inclined to look for in every language. It seems desirable to show here in how far the system with which we are familiar is characteristic only of certain groups of languages, and in how far other systems may be substituted for it.

The structural linguists came to the conclusion that languages needed to be analyzed within their own systems, in terms of their own grammatical structures, instead of being forced to fit into pre-established grammatical systems based on different grammatical structures. In order to achieve this goal, they devoted themselves to collecting data and to observing and describing many languages as accurately as possible, instead of making generalizations about the languages.

Bloomfield (1935) also criticized the tendency of earlier English grammarians to blindly
accept the grammar established by Greek and Latin scholars, and he warned against making generalizations based on studies done on a limited number of languages, especially similar languages. He wrote:

[A] knowledge of foreign types of grammatical structure…would have opened their eyes to the fact that even the fundamental features of Indo-European grammar, such as, especially, the part-of-speech system, are by no means universal in human speech. Believing these features to be universal, they resorted, whenever they dealt with fundamentals, to philosophical and psychological pseudo-explanations. (Bloomfield 1935:17)

He continued,

It is a mistake to suppose that our part-of-speech system represents universal features of human expression. If such classes as objects, actions, and qualities exist apart from our language, as realities either of physics or of human psychology, then, of course, they exist all over the world, but it would still be true that many languages lack corresponding parts of speech. (Bloomfield 1935:198–199)

Thus, the structural linguists raised two important issues for parts of speech. First, they suggested that word classification systems can vary depending upon the language, and hence the traditional part-of-speech system is not appropriate to describe different types of languages. Second, for the same reason, they rejected the universality of parts of speech.

As mentioned earlier, the structuralists collected data on languages that were new to
them. They elicited speech samples from native speakers, and documented the samples precisely to form a corpus. Then they analyzed the data by first classifying the elements of the language, beginning with the smallest unit, the sounds, and then moving on to words, and finally, to the largest unit, sentences. What was important for the structural linguists in classifying language elements was the environment in which elements occur. For instance, if two morphemes occur in the same environment, then they belong to the same morpheme class. The structuralists did not separate syntax and morphology; their categories list affixes and words together. For instance, using the terminology from the traditional part-of-speech system and adopting the distributional method, Harris (1951) identified morpheme classes with morpheme class markers for English: the morpheme class marker -ly identifies the morpheme large as class A; the ___ or the large ___ can identify the class N, such as man, auto, and life; -ing identifies the class V, such as do, have, see, and so on (Harris 1951:255–256). N(ouns) can be differentiated from affixes, because an N is a constituent of a sentence, while an affix is a constituent of a word. Harris used the notations V(erb), which is also subdivided into smaller classes using distributional methods, and A(djective). He also came up with new notations such as D (very, well), R (do, will), and T (a, some). Other linguists followed structuralist principles, and avoided using the traditional part-of-speech terminology. Fries (1952) classified words in English using simple labels, Class 1 to Class 4, for the “parts of speech” or “form class” and Group A, Group B, and so on, for “function words.” His classification relied on test frames: words that can replace each other in a given position form the same category. Fries’s list does not include affixes, unlike Harris’s. Therefore, his categories are syntactic categories, determined solely by distribution. By avoiding using the terms “noun” and “verb,” yet calling his word classes “parts of speech,” he showed that word classes could be determined by the structure of each language under investigation.
In sum, the structural linguists questioned the validity of parts of speech, especially as a universal feature, and opened up a new approach to word classification, which was language-specific. At the same time, their approach introduced a new view of parts of speech as syntactic categories.

2.4. Categories as Universal

The opposite approach to word categories is taken by a school of linguistics that shares a fundamental view of language and the study of languages with Cartesian philosophy: Language is the representation of human minds, and the study of languages is the investigation of the relationship of language and mind. Noam Chomsky found the structuralists’ approaches to the study of language problematic. First, he was against the inductive method of working on a corpus (i.e., a finite set of utterances) to understand the grammar of a language under investigation. He pointed out that, judging from the fact that children acquire language despite the fact that the information a child receives is not always complete and grammatical, the human mind is capable of learning the grammar of a language without overt instruction. Furthermore, children learn a language’s grammar in spite of the fact that every child is exposed to a limited and different set of information. Chomsky claimed that the human brain must, therefore, be genetically hard-wired for learning language, and equipped with an innate language faculty. He thus defined linguistics as the search for the innate grammar, Universal Grammar (UG), which would include the properties that are shared by all human languages. Generative grammar is not concerned with the rules of individual languages, but aims at building a set of rules that will generate well-formed sentences and that applies to all human languages. In other words, I-language, which is the internal linguistic knowledge in the mind of the speaker, is the object of
study in linguistics, not E-language, which is the actual linguistic output, such as sentences. A corpus that is a collection of sample utterances by native speakers (E-language) does not always reflect the I-language. Chomsky argued that the linguists’ task should not be merely describing a language based on a corpus, which includes utterances that do not comply with rules of well-formedness. Grammar that adequately describes and explains I-language and that generates well-formed sentences for all human languages is abstract in nature.

This model of grammar uses representations of sentences with a top-down structure, as opposed to the bottom-up approach of the structuralists, whose analysis moved from the smallest unit (phonetic representation) to the largest (sentence). Chomsky posited, in his early grammatical model, sets of rules that are responsible for forming grammatical sentences (phrase-structure rules), moving phrases into the proper places (transformational rules), and changing words to the proper forms (morphophonemic rules). Thus, his grammar was componential, that is, the distinct components of grammar are assumed to work separately. To describe such rules, Chomsky used simple notations such as S(entence) → NP + VP and VP → Verb + NP, in which NP and VP mean Noun Phrase and Verb Phrase. Thus, categories such as N and V are primarily concerned with syntax. The syntactic categories were further divided into two groups: lexical categories (e.g., N, V) and functional categories (e.g., Determiner, Auxiliary). In the 1965 model, the idea of “lexicon” was introduced as one of the base components of grammar, which worked together with the phrase-structure rules to produce deep structure, an abstract sentence structure. The transformational rules applied to the deep structure to produce a final well-formed sentence, namely, surface structure. The lexicon is a list of words that includes both lexical and

---

3 It must be noted here that lexical categories can mean two things: categories that belong to lexical items, or lexical categories versus functional categories. In this dissertation, lexical categories will mean the former, unless otherwise specified.
functional categories. The lexicon stores the information necessary for the successful derivation of a sentence, and, unlike the earlier model, which was exclusively a syntactic model, the newer model included semantic concepts. Items in the lexicon contain necessary information including category, subcategorization properties, semantic features such as animate/inanimate, and inflectional and derivational morphemes. The words are prepared to be syntax-ready in the lexicon. In both models, despite their differences (i.e., the later inclusion of semantic factors in the items in the lexicon), the categories used in this grammar are considered universal. What N and V signified is not explained in either model, as the major categories used in Chomskyan grammars—Noun, Verb, Adjective, and such—are assumed to exist in all languages. This is the logical conclusion that follows the premises of Chomskyan linguistics that a universal grammar can explain how the human mind operates in terms of language. Therefore, the categories used in grammar are hypothetical in nature. Chomsky explained his understanding of the categories when he wrote that their definitions depended on “the assumption that the symbols S, NP, VP, N and V have been characterized as grammatical universals” (Chomsky 1965:72) and that “there is no reason to rule out, a priori, the traditional view that such substantive characterizations must ultimately refer to semantic concepts of one sort or another” (Chomsky1965:117). Note that, in this syntactic theory, the categories are assumed to be universal and what characterizes the categories most fundamentally is meaning.

2.5. Confusion Regarding Categories

Looking at Chomsky’s 1965 model, illustrated in Figure 2.1 below, it is notable that what structuralists such as Harris and Fries called “categories” are not the same as what Chomsky refers to as categories.
The structuralists’ investigations were based on corpora (E-language), which belong at the level of surface structure in the above model. According to structuralist principles, different languages have their own categories based on their own grammatical structures. Consequently, the categories the structuralists refer to are language-specific. Meanwhile, in generative grammar, the search for UG assumes that the categories are linguistic primitives, universal, and inherent to the lexical items.

In sum, to revisit the origin of the notion of parts of speech, it began with the word classes identified for Greek, and was adopted and developed by and for Greek, Latin, and other Romance languages; therefore, the part-of-speech concept is language-specific. The study of languages in the past was motivated by two different views of language. On the one hand were scholars who documented and described the works of great writers in order to preserve them for
future generations, and on the other hand were scholars who viewed language as the expression of thoughts and sought to discover the relationship between language and the human mind. For both, however, the part-of-speech system was based on morphology and/or semantics. A similar difference in the motivation for the study of language exists between the structuralists and the Chomskyan linguists. The structuralists are similar to the Greek philologists in the sense that their goal is the compilation of language-specific data, while Chomskyan linguists, influenced by Cartesian philosophy, echo the Greek philosophers in their search for the universal features of language. However, unlike the morphology-based word classification system of the past, both structural grammar and Chomskyan grammar use distributional methods for categorizing words; that is, the categories they refer to are syntactic categories. Meanwhile, the use of terms meaning “parts of speech,” “noun,” “verb,” and such, has for the most part remained unchanged. Table 2.1 below presents a summary of this information.
Table 2.1. Parts of speech in the different grammatical models

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Use of “noun, verb,” etc.</th>
<th>Criteria</th>
<th>Universal or language-specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greek/Latin grammars</td>
<td>parts of speech</td>
<td>Yes</td>
<td>Morphology, language-specific</td>
</tr>
<tr>
<td><em>Port-Royal Grammar</em></td>
<td>parts of speech</td>
<td>Yes</td>
<td>Semantics, aims at universal</td>
</tr>
<tr>
<td>Structuralists</td>
<td>word class, form class, morpheme class</td>
<td>yes/no</td>
<td>Syntax, language-specific</td>
</tr>
<tr>
<td>(Early) Chomskyan</td>
<td>Category</td>
<td>Yes</td>
<td>Syntax, Universal</td>
</tr>
</tbody>
</table>

Note. Early Chomskyan: Standard theory

2.6. Categories as Features

In the previous section, I explained that the early Chomskyan models of grammar (Standard Theory and Extended Standard Theory) showed a new direction for the study of languages in the sense that their objective was to search for the relationship between language and the human mind, not to document and describe individual languages. A generative grammar should describe and explain how we successfully form grammatical utterances in our minds and how we learn the process for doing so, and it should also generate well-formed utterances. Based on its theoretical assumption of universal properties for all languages, such a grammar should apply to all natural languages and account for their linguistic phenomena, regardless of superficial differences such as word order, word formation strategies, and so forth. In this view, languages have a deep underlying linguistic representation that is abstract in nature, consisting of the rules to build and rewrite well-formed utterances, which interact with the lexicon, where
lexical items are stored with the necessary information. The rules in the early Chomskyan models often resembled mathematical formulae, with arrows, brackets, and code-like notations, and sentence structure was schematically described in a top-down manner in tree diagrams. The categories were considered to be universal and abstract, and were postulated a priori. Chomsky used category notations such as N and V, and the categories were described by features in a binary system. For example, Chomsky (1965:68) described the sentence *Sincerity may frighten the boy* using the following notations:

(2.1)  
\[
S \rightarrow NP \rightarrow Aux \rightarrow VP \\
\hspace{1cm} VP \rightarrow V \leftarrow NP \\
\hspace{1cm} NP \rightarrow Det \leftarrow N \\
\hspace{1cm} NP \rightarrow N \\
\hspace{1cm} Det \rightarrow the \\
\hspace{1cm} Aux \rightarrow M \\
\hspace{1cm} M \rightarrow may \\
\hspace{1cm} N \rightarrow sincerity \\
\hspace{1cm} N \rightarrow boy \\
\hspace{1cm} V \rightarrow frighten
\]

He then gave the following definitions: “A category that appears on the left in a lexical rule we shall call a *lexical category*; a lexical category or a category that dominates a string…X…, where X is a lexical category, we shall call a *major category*” (Chomsky 1965:74). He first said N, V, and M are lexical categories in (2.1) and that “all categories except Det (and possibly M and Aux) are *major categories*” (Chomsky 1965:74). The actual words, *lexical formatives* in Chomsky’s term, were assigned features, such as [+Common] or [+Transitive], which were called *syntactic features*. Thus, the lexical categories were also described with features. For example, the category N was described with syntactic features as follows (Chomsky 1965:85):
(2.2)  $N \rightarrow [+N, \pm \text{Common}]$

$[+\text{Common}] \rightarrow [\pm \text{Count}]$

$[+\text{Count}] \rightarrow [\pm \text{Animate}]$

$[-\text{Common}] \rightarrow [\pm \text{Animate}]$

$[+\text{Animate}] \rightarrow [\pm \text{Human}]$

$[-\text{Count}] \rightarrow [\pm \text{Abstract}]$

In generative grammar, complex features are a part of each lexical item stored in the lexicon, and these features interact with the different principles during the course of derivation.

For the lexical categories, notations $N$, $V$, $A$, and $P$ are used, but they do not refer to the traditional parts of speech, especially $A$ and $P$. The functional categories, including the newly created Agreemen, Tense, and so forth, are also defined in terms of features. The lexicon includes both lexical items with inherent categories and functional items, which are inserted for syntactic derivations. Although the lexical categories $N$, $V$, $A$, and $P$ are assumed to be universal, this does not mean that all languages have all the categories. In addition, how a language sorts its lexical items into the different categories of $N$, $V$, and so on is language-specific.

Unlike the structuralists’ bottom-up approach, in which morphemes combine to form words, and words combine to form clauses and phrases and ultimately sentences, generative grammar is concerned with the well-formedness of a sentence in terms of the distribution and the configuration of the lexical items that are combined.
Figure 2.2. Tree structure representation in generative grammar

Syntactic rules are responsible for checking the compatibility of the lexical items that fill the terminal nodes in the tree above. For example, the functional category T, which is the head of the TP, must take a VP complement; V, as the head of the VP, must take DP for its complement. However, syntax is not concerned with the properties of N, V, A, and P. The lexical categories are an inherent part of the lexical items in the lexicon, and the syntactic rules operate on the lexical items that are readied for derivation in the lexicon. Therefore, in generative syntactic theory, the only possible way to define a lexical category is in terms of the lexical category X being the head of XP. As this definition suffices for the framework, there has been little
discussion regarding syntactic categories among generativists, except Baker (2003). Working in the framework of the principles and parameters theory, Baker (2003:21) attempts to give a more substantial definition than notations such as “Noun is [+N, -V], verb is [-N, +V]

\[(2.3) \quad \text{Noun is } +N = \text{“has a referential index”} \]
\[(2.3) \quad \text{Verb is } +V = \text{“has a specifier”} \]
\[(2.3) \quad \text{Adjective is } -N, -V \]
\[(2.3) \quad \text{Preposition is part of a different system (functional)} \]

Baker characterizes a verb as having a specifier, a syntactic position that receives a theta-role of agent of theme from the verb, which makes verbs licensors of the subject. Although this definition is based on the syntactic spec-head configuration, Baker believes that the syntactic definition can also explain the semantic properties. As verbs license the subject in the specifier position, they are predicates that denote action or events. Therefore, Baker claims that syntax-centered definitions of lexical categories can address the semantic properties of the lexical categories. He also claims that the category bearers are not lexical items in the lexicon, but that categories are determined by syntax; a local configuration, to be precise. This syntax-oriented approach that rejects categories being a property of lexical items will be discussed in Chapter 4.

Generative grammar is a syntax-oriented theory, for which the lexical categories are syntactic primitives on which the syntactic rules can operate. The categories used in the theory include not only the lexical categories, but also the functional categories. The traditional parts of speech N, V, A, and P were redefined syntactically, that is, in terms of distribution and configuration. As the syntactic structures and rules in generative grammar are universal, the categories, N, V, A, and P are universal as well, which means they exist in all human languages. However, how each language classifies its lexical items into these classes or whether a language possesses all four classes is a language-specific issue, and such issues are not in the domain of syntax. In this approach, a lexical item with multiple functions is stored in the lexicon with
separate entries (N: change; V: change), or one entry with multiple uses, one being derived from
the other (V: change → N: change), or one without a preassigned category that receives its
category from syntax (in a local configuration). Generative syntax does not concern itself with
the processes and mechanisms of word categorization.

2.7. Categories as Prototypes

In generative grammar, the categories of concern are syntactic categories, which are
postulated a priori and are considered universal, and are described with syntactic features using
binary feature notation. There have been opposing views. In this section, I will summarize
different approaches to word categorization.

2.7.1. Against syntax-based word classification

By the late 1960s, linguists were already voicing concerns with word classification that
relied on syntactic behavior. Generative semanticists such as George Lakoff (1970) and James
McCawley (1973) maintained the concepts of UG, the derivational aspect of the generative
approach to grammar, and, in the beginning, the level of deep structure. They later claimed that
there was an underlying semantic structure from which syntactic structure was derived, rather
than Chomsky’s syntactic deep structure. In their theory, the underlying level had only three
universal and abstract categories—Sentence, Noun, and Verb—and language-specific
transformational rules were applied to derive the surface structure categories. This was their
solution to a problem of early Chomskyan grammar, according to which the categories in surface
and deep structure should be the same. Generative semantics never demonstrated how the surface
structure could be derived from the abstract semantic-oriented underlying structure by
transformational rules that would be responsible even for affixation processes, and thus the
theory fell out of favor. However, the objection to syntacticocentric grammatical theories remained.

The crucial role that semantics plays in grammar was also emphasized by Charles J. Fillmore (1968), who first proposed in the 1960s that syntactic structure could be predicted by the semantics of the lexical items. Like the generative semanticists, Fillmore believed that syntax and semantics were not separate, and that it was the semantics that determined deep structure. He introduced case grammar, in which the semantics of verbs plays the central role in sentence structure. “Deep cases,” such as Agentive and Dative, which are specified by the verb, represent the relationship between syntax and semantics. For instance, the semantic requirement of the English verb *give* that the act of *give* has the semantic requirement of having an Agent, Object, and Beneficiary determines the syntactic structure as ditransitive. Case grammar assumed that verbs form “case frames,” determined semantically by the number of deep cases they select. Like generative semantics, case grammar posits deep underlying semantic structures to which various transformational rules apply to derive the surface structure. Moreover, in both generative semantics and case grammar, the nature of the categories was not clear: are they semantic or syntactic? Although generative semantics and case grammar did not survive as mainstream theories, they laid a foundation for the development of cognitive grammar and construction grammar.

### 2.7.2. Language as a part of cognition

The concept of parts of speech emerged from the study of ancient Greek, which had rich morphological inflection. The part-of-speech system was primarily morphology-based, although

---

4 Case grammar later developed into frame semantic theory.
it considered semantics when appropriate. For the scholars who wrote the *Port-Royal Grammar*, not word forms but word meanings had the central role in the study of language. Structural linguists, upon encountering so-called exotic languages, noticed that the traditional classification of parts of speech did not work for these non-Indo-European languages, and thus devoted themselves to the task of documenting and describing the languages without imposing pre-existing measures. In order to achieve their goal (i.e., to analyze and describe the previously undescribed languages), they observed word distribution, therefore relying on syntax, to determine word classes. The cognitive linguistics movement, pioneered by Ronald Langacker in the 1980s, rejected Chomsky’s theory in its core principles. Cognitive grammar does not assume linguistic competence and universal grammar as an innate faculty. Rather, language is considered to be a part of human cognition, not as an autonomous entity, but as an integral part of it. The speaker’s language acquisition is usage-based, not a product of innately given linguistic ability. Cognitive grammar does not assume a deep underlying structure to which transformation rules apply to derive a surface structure. The central elements in this approach are symbols, which are composed of semantic and phonological units as the basic units of language. Small symbolic units such as affixes and inflectional morphemes combine with word units, which in turn combine with other units to make larger units, such as phrases, clauses, and sentences. Grammar being a part of our general cognitive ability means that grammar is a conceptual representation; this means that not only meanings but also sounds, morphemes, and sentence structures are all conceptual, in the sense that they all have to be comprehended and processed in our minds. In the absence of innately given language acquisition ability, it is through everyday experience that speakers acquire linguistic knowledge, which is a mentally represented inventory of well-formed and accepted units. Cognitive grammar is thus a grammar of conceptual representations of our
cognitive ability in language use, and the categories used in this approach reflect this philosophy. Before cognitive grammar, the categories in what the cognitive grammarians call the “classic model” were clearly defined; rigid boundaries separate the items that belong to different groups. In other words, an item must be either inside or outside of a given group. In the real world, however, categorization is not always clear-cut, and there are always cases where we cannot easily decide the category of an item in question. Psychologists Eleanor Rosch and Carolyn Mervis drew on Wittgenstein’s discussion in his 1953 book, *Philosophical Investigations*, of “family resemblance” as a factor for grouping items together in the same category. The items in a family resemblance relationship are categorized together not by one common element shared by all the items, but one or more common element(s) that the items share with others in the relationship. In other words, the items show overlapping similarities, rather than one common feature. This principle was adopted by Rosch (1973) and Rosch and Mervis (1975), who proposed that cognitive categories are formed around a prototype, that is, the best example of the category. Cognitive grammar adopts this approach to grammatical categories, focusing on the prototypes of a category rather than defining it by a set of properties. Members that belong to the same category show degrees of membership: Some members possess high relevancy to the central features that best represent the category, while others show low relevance. According to this model, the boundaries that separate categories are “fuzzy,” unlike in the classic model, which had sharp demarcation between categories. This prototype approach was elaborated by Croft (1991, 2001, 2004) in his theory of radical construction grammar.

Construction grammar, like cognitive grammar, was developed in the 1980s by linguists such as Charles Fillmore (1968), Paul Postal (1964), and George Lakoff (1970) who did not agree with syntax-oriented generative grammar. Some common principles are shared by these
two approaches, including meaning being the central part of grammar, grammar being considered part of human cognition, and grammar being usage-based. There are different models within construction grammar, but they all take the construction to be the basic unit of grammar. A construction is a conceptual/symbolic unit, which is form and meaning linked together, like a lexical item, and a construction as a whole, whether it is a word or a sentence, is treated as a sign. Construction grammar, like cognitive grammar, does not assume that grammar is componential. It aims at providing a uniform model of grammar that represents not only syntactic but also semantic and morphological rules. Unlike the generative approach in which lexical items are the minimal syntactic units, constructions can be atomic (word/lexical item) or complex (phrase, sentence). This interpretation of constructions suggests that there is a continuum between the lexicon and syntactic constructions. The following table shows the syntax–lexicon continuum (Croft 2001:17):

Table 2.2. The syntax-lexicon continuum

<table>
<thead>
<tr>
<th>Construction type</th>
<th>Traditional name</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex and (mostly) schematic</td>
<td>syntax</td>
<td>[SBJ be-TNS Verb-en by OBL]</td>
</tr>
<tr>
<td>Complex and (mostly) specific</td>
<td>idiom</td>
<td>[pull-TNS NP-'s leg]</td>
</tr>
<tr>
<td>Complex but bound</td>
<td>morphology</td>
<td>[NOUN-s], [VERB-TNS]</td>
</tr>
<tr>
<td>Atomic and schematic</td>
<td>syntactic category</td>
<td>[DEM], [ADJ]</td>
</tr>
<tr>
<td>Atomic and specific</td>
<td>word/lexicon</td>
<td>[this], [green]</td>
</tr>
</tbody>
</table>

As we can see in the table, constructions can be as small as an individual lexical item or as large as a sentence, and the components of grammar do not exist in isolation; they form a continuum.

Croft (1990, 2001, 2004) took constructions, not just lexical items, as the object of analysis for categories. He proposed a prototype approach in which various types of constructions are classified in terms of their function (reference, modification, or predication)
and the concept they denote (object, property, or action). The approach is summarized in table 2.3 (Croft 1990:187):

**Table 2.3. Conceptual space for parts of speech**

<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
<th>Modification</th>
<th>Predication</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBJECT</td>
<td>object reference</td>
<td>object modifier</td>
<td>object predication identity predication</td>
</tr>
<tr>
<td>PROPERTIES</td>
<td>property reference</td>
<td>property modifier</td>
<td>property predication location predication</td>
</tr>
<tr>
<td>ACTIONS</td>
<td>action reference</td>
<td>action modifier</td>
<td>action predication</td>
</tr>
</tbody>
</table>

As the table shows, Croft uses a conceptual space that maps functions and concepts to identify the prototypes of the categories.

**Table 2.4. Prototypical parts of speech**

<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
<th>Modification</th>
<th>Predication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objects</td>
<td>UNMARKED NOUNS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Properties</td>
<td>UNMARKED ADJECTIVES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actions</td>
<td></td>
<td></td>
<td>UNMARKED VERBS</td>
</tr>
</tbody>
</table>

The constructions that are UNMARKED are the prototypical categories, exhibiting the best examples of the membership. UNMARKED NOUNS, for example, are used for one of the three universal propositional acts, reference, and used for the prototypical function, to denote objects. The prototype constructions need zero structural coding, and thus are UNMARKED, which means they are not accompanied by any derivational morphemes or prepositional phrases; that is,
they fulfill their function without further measures being taken. Therefore, according to this
model, book, for instance, is a prototypical noun, and its adjective form, bookish, is not a
prototypical adjective, because a further measure was taken—the use of the derivational suffix -
ish—for it to be used as a modifier. Jump is a prototypical verb, as it denotes an action, and thus
requires -ing (gerund) or needs zero-derivation to become an action nominal jump (n) to be used
for reference. Table 2.5 presents types of non-prototype category constructions (Croft 1991:
185):

**Table 2.5. Overtly marked structural coding constructions for parts of speech**

<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
<th>Modification</th>
<th>Predication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objects</td>
<td>UNMARKED NOUNS</td>
<td>genitives, adjectivalizations, PPs on nouns</td>
<td>predicate nominals, copulas</td>
</tr>
<tr>
<td>Properties</td>
<td>deadjectival nouns</td>
<td>UNMARKED ADJECTIVES</td>
<td>predicate adjectives, copulas</td>
</tr>
<tr>
<td>Actions</td>
<td>action nominals, complements, infinitives, gerunds</td>
<td>participles, relative clauses</td>
<td>UNMARKED VERBS</td>
</tr>
</tbody>
</table>

The prototype approach describes parts of speech by mapping form, function, and concept so that
the element in question can be located somewhere in the table. Croft finds the domain for each
part of speech in its “typological pattern in conceptual space” (2001:65) meaning that such terms
as “noun, verb, and adjective” are functional prototypes, which describe the core of a category.
They are not categories that exist only in some languages; rather, as language universals, they
exist as typological patterns and tendencies in all languages. What Croft accomplished in this
analysis was to untangle the confusion that often accompanies the discussion of parts of speech.
His map includes all three criteria used to identify the parts of speech: the semantic criterion in
the concepts of objects, properties, and actions; the syntactic criterion in the concept of the propositional acts of reference, modification, and predication; and the morphological/syntactic criterion in the concept of “the further measures” that fill all the non-prototype spaces in the map. He shows that the prototypical categories must satisfy all three criteria, while the non-prototypical members do not.

2.8. Parts of Speech in Typology

Linguistic typology is concerned with cross-linguistic comparison and the classification of languages. Typologists often focus on one particular linguistic phenomenon, such as word order, and classify the languages they investigate into different types based on the languages’ realization of the phenomenon (e.g., SOV or SVO languages). Based on their findings, typologists seek to make generalizations, such as Greenberg’s (1963) implicational universals, in which he found that languages with SOV word order tend to have postpositions. In the previous section, I introduced Croft’s (1991, 2001, 2004) prototype analysis of parts of speech. What Croft showed is that, typologically speaking, there will never be a language that possesses prototypical members of parts of speech (noun, verb, adjective) with marked constructions, whereas he implies that it is possible that there are languages that do not have overt markers for non-prototypical members of parts of speech. In fact, there are many such examples: for instance, action-denoting nouns such as *hit*, *cut*, and *mail* in English.

Parts of speech are a very important grammatical issue for typology, as one can see from the fact that the sign V for Verb in SVO is used to classify languages into types by the word order. Some typologists have addressed the issue that some languages do not appear to have morphological part-of-speech distinctions. Hengeveld (1992) and Hengeveld, Rijkhoff, and
Siewierska (2004), from typologist and functionalist points of view, take samples from diverse languages and compare their different types of part-of-speech distinctions, based on the functions of lexical items. Using the traditional part-of-speech system and terminology, Hengeveld (1992) first divides the major parts of speech into four (noun, verb, adjective, and manner adverb) based on the four syntactic slots a word can occupy. Noun and adjective are the head and modifier of a referential phrase respectively, and verb and manner adverb are the head and modifier of a predicate phrase:

(2.4) V is a lexeme that can be used as the head of a predicate phrase.  
N is a lexeme that can be used as the head of a referential phrase.  
A is a lexeme that can be used as a modifier within a referential phrase.  
MAdv (Manner Adverb) is a lexeme that can be used as a modifier within a predicate phrase. (Hengeveld 2013: 33)

He defines the four categories further as follows:5

(2.5) A *verbal* predicate is a predicate which, without further measures being taken, has a predicative use *only*.  
A *nominal* predicate is a predicate which, without further measures being taken, can be used as the head of a term.6  
An adjective predicate is a predicate which, without further measures being taken, can be used as a modifier of a nominal head.  
An adverbial predicate is a predicate which, without further measures being taken, can be used as a modifier of a non-nominal head. (1992:37)

Hengeveld proposes that the languages he investigated can be classified into three groups according to their part-of-speech systems: flexible, differentiated, and rigid.

---

5 Hengeveld explains that the term “predicates” in these definitions refers to “lexemes.”  
6 “Term” refers to a referential expression with a nominal head. A term is inserted in the argument position in the predicate.
Table 2.6. Hengeveld’s parts-of-speech systems

<table>
<thead>
<tr>
<th>PoS-system</th>
<th>Head of predicate phrase</th>
<th>Head of referential phrase</th>
<th>Modifier of head of referential phrase</th>
<th>Modifier of head of predicate phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexible</td>
<td>1</td>
<td>Contentive</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Verb</td>
<td>non-verb</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Verb</td>
<td>Noun</td>
<td>Modifier</td>
</tr>
<tr>
<td>Differentiated</td>
<td>4</td>
<td>Verb</td>
<td>Noun</td>
<td>Adjective</td>
</tr>
<tr>
<td>Rigid</td>
<td>5</td>
<td>Verb</td>
<td>Noun</td>
<td>Adjective</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Verb</td>
<td>Noun</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Verb</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Hengeveld 2013:36)

According to the table above, languages such as English that possess all four categories have a differentiated system. If a language allows the same item to be used in two or more slots, this language has a flexible system. For instance, in Māori, *pai* can be used as a modifier as in *te tangata pai* ‘the good man’, and also in the predicate as in *ka pai* ‘(it is) good’. *Pai* is used in both functions without changing its word form. This implies that verb and adjective classes are lumped together into one. In rigid languages, on the other hand, a non-lexical strategy must be used to compensate for the lack of a particular category. For example, if a language does not have a class of words that can be used as modifiers in a referential phrase such as *big*, a noun can compensate for the absence of a modifier class and be used for this purpose as part of a prepositional phrase modifier as in *a man of bigness*. Hengeveld (2013) considers that rigidity and flexibility are scalar, and suggests that the “extreme” flexible languages make no lexical distinction among the major categories, and include many multifunctional lexical items.
Hengeveld et al. (2004) use the term “contentive” to refer to the lumped category N/V/A/Adv. (Note that Hengeveld and his colleagues discuss traditional part-of-speech systems and use the traditional terms, while they define them syntactically.) What “contentive” appears to imply is that there are languages that do not have word class differentiation.

Evans and Osada (2005) warn that vigorous criteria have to be satisfied in order to determine that a given language really does not have word class differentiation. They offer three criteria for establishing lack of word class distinctions: bidirectionality, compositionality, and exhaustiveness. The first criterion means that the members of the merged classes show identical privileges, which means that within the merged classes, the lexical items are freely exchangeable in the two syntactic positions. The following examples from Mundari are from Evans and Osada (2005:354–355).

(2.6) buru=ko bai-de-d-a
      mountain=3PL.S make-COMPL-TR-INDIC
     ‘They made the mountain.’

(2.7) saan=ko buru-ke-d-a
      firewood=3pl.S “mountain”-COMPL-TR-INDIC
     ‘They heaped up the firewood.’

In the above examples, buru ‘mountain’ can combine with the subject clitic =ko as well as the verbal affixes, -ke-d-a. This means that buru can be in both argument phrase and predicate phrase.

The second criterion, compositionality, means that the difference in the meaning of the words in two different syntactic positions must be predictable from the function of the syntactic positions. For example, one can predict that dance in VP means the predicate ‘dances, is dancing’ while dancer in NP means the argument ‘the one who dances, or is dancing’, which is a reference. The third criterion, exhaustiveness, means that in order to claim that a given language
has flexible word class, the first and second criteria need to apply to all relevant words in the lexicon of the language. Evans and Osada argue that, when these criteria are applied, prior analyses that claimed to find flexible word classes are inadequate, and that they have generally relied on very limited data that led to the misleading conclusion that the languages under investigation were monocategorial or lacked a noun and verb distinction.

Despite disagreement on the flexible category, researchers in the generation of generative grammar and later usually assume that parts of speech are universal and retain the notations N, V, and A. Some typologists reject the idea of applying pre-existing knowledge from much-studied languages to less-studied languages. Gil (2000, 2013) and Broschart (1997), for instance, share the same basic idea, according to which the grammatical analysis of a language must be done within its own systems, instead of being forced to fit pre-established grammatical systems based on different grammatical structures. As Haspelmath (2007) claims, structural categories are language-particular, as different languages have different grammatical structures. Therefore, these scholars share a basic tenet of structuralism.

Gil (2000, 2013) claims that the word categories of a language under investigation can be determined only after a comprehensive description of syntactic behaviors is completed, and for that purpose, the part-of-speech system, which itself is a product of Eurocentric analysis, cannot be applied to reach a correct analysis of non-European languages. Broschart (1997) also proposes that word classes of some languages cannot be classified by the traditional part-of-speech system. He shows that Tongan has its own categories and does not have the noun/verb distinction. Because Tongan does have predicative phrases, preceded by a tense/aspect marker and referential phrases, preceded by an article, some might suggest that Tongan lacks lexical categories. However, according to Broschart, the matter is simply that, if there are no lexical
nouns and verbs, it is impossible to posit nominal and verbal syntax. He denies the noun/verb distinction in Tongan on both lexical and syntactic levels. Consequently, he seeks and offers an alternate account of Tongan grammar that does not have to refer to nouns and verbs. These approaches are discussed in detail in Chapter 4.

As more research has been done on languages with flexible word classes, the implications of the typological findings for word categories have also been researched. Hengeveld (1992, 2013) and Hengeveld et al. (2004) propose the part of speech hierarchy:

\[
\text{(Head of pred. phrase)} > \text{Head of ref. phrase} > \text{Modifier of head of ref. phrase} > \text{Modifier of head of pred. phrase}
\]

According to this hierarchy, more languages tend to have the categories on the left than the ones on the right, and languages that have the category of “modifier of head of reference phrase” also have the other two categories on the left. According to Van Lier and Rijkhoff (2013), in an unpublished manuscript written in 2006, Geoffrey Haig proposes the “principle of successively increasing categorization,” by which he means that the categorial distinctions will increase with the level of grammatical complexity. In other words, less categorial distinction will occur on the lower levels of grammatical complexity, such as with roots and inputs of derivation, and more distinction will occur on the higher levels, such as outputs of derivational and inflectional morphology. In addition, Frajzyngier and Shay (2003) suggest that, if one grammatical level has flexibility in category specification, then it has to be counterbalanced at another level. Therefore, if a lexical item lacks inherent lexical category, the category specification has to be compensated for by other means such as affixations, adpositions, and linear order.

The typologists’ main concern is the variation of part-of-speech systems among languages, mainly from the empirical point of view, and they seek generalizations—what the patterns and the types of certain linguistic phenomena imply for other areas of grammar. As
more relationships are found between the part-of-speech system and other grammatical phenomena, such as correlations between degree of flexibility and the types of lexical categories a language possesses, part-of-speech systems may become an important determinant in the study of typology.

2.9. Conclusion

In this chapter, I gave a brief overview of the parts of speech in the study of language. The classification of words into different parts of speech was initiated by the Greek philologists and philosophers. Their classification was morphology-oriented, based on their language’s inflectional system, and notional definitions were given when necessary. The *Port-Royal Grammar* kept the part-of-speech system of antiquity, but defined the word categories in terms of semantic instead of morphological criteria. The American structuralists realized that this traditional word classification system developed by and for Indo-European languages failed to accurately describe some non-Indo-European languages. The structural linguists therefore tried to describe hitherto undocumented languages without imposing pre-existing grammatical models. Some avoided using the terms noun, verb, adjective, and so forth, while others kept this terminology with different definitions, but the word classification used by the structural linguists was not based on morphology or semantics; it was a syntactic notion defined by the distributional properties of words or morphemes. Since the American structuralist era, nouns, verbs, and such have been taken as syntactic categories. Generative grammar, led by Chomsky, is a syntactic theory, and the grammatical categories necessary for the syntactic models are syntactic categories. The syntactic categories, N, V, A, and so on, are treated as syntactic features inherent to lexical items. Cognitive grammar uses semantic criteria (which is also
implied in the Greek philology model), recognizes fuzzy boundaries, and deals with this problem by distinguishing unmarked (root) versus derived words. Hengeveld addresses the issue of morphologically unmarked items appearing in non-prototypical slots by suggesting rigid versus flexible boundary systems. Gil and Broschart propose that some languages with flexible systems, in Hengeveld’s term, do not have lexical categories. Broschart in particular proposes an alternative classification using a different set of criteria. The approaches that assume that lexical items do not have category will be discussed in chapter 4.

To conclude, three main points have emerged from this chapter’s discussion: (1) “parts of speech” is originally a Greek-based, language-specific word classification, primarily based on morphological criteria with semantic criteria as secondary; (2) “parts of speech,” or at least the terminology of N(noun), V(erb), and such, is often used to mean syntactic category; and (3) it is not clear if “parts of speech” or noun/verb/adjective distinctions exist in all languages. What needs to be cleared up first is the terminology of parts of speech and syntactic categories. The original parts of speech were defined morphologically and semantically. American structuralists used distributional criteria for word classification, and syntax became the center of linguistic study with the advent of generative grammar; since that time, the parts of speech and syntactic categories have been used almost synonymously, which seems contradictory. These two terms should not be identified as the same linguistic phenomenon. Parts of speech can be defined morphologically, notionally, and syntactically, while syntactic categories, on the other hand, are the categories assigned to the grammatical units of syntactic structure. Further, a distinction between lexical and syntactic categories should be made, as illustrated in Broschart’s analysis. Finally, do all languages have at least N(oun) and V(erb)? If nouns are typically used for referential phrases and verbs are typically used for predicative phrase, as Hengeveld (1992, 2013
and Croft (1990, 2001, 2004) claim, then it is reasonable to assume that all human languages have nouns and verbs, an assumption with which Broschart (1997) and Gil (2000, 2013) disagree. Māori is known to have fluid word categories, showing overlapping of classes. In the next chapter, examples of Māori will be introduced and how Māori grammarians have dealt with word classes will be discussed.
3.1. Introduction

As chapter 2 explained, the idea of parts of speech was developed by scholars in ancient Greece and inherited by scholars who spoke and studied other Indo-European languages. In these languages, inflectional morphemes help identify the categories, and the notional definitions support the class identification. The syntactic definition for parts of speech became the norm when the American structuralists began investigating non-European languages that had not been described before. For languages such as Māori that do not have rich inflectional systems, linguists cannot rely on morphological criteria for word classification, and distributional criteria often produce a high degree of category overlapping. In this chapter, I will give a brief grammatical overview of Māori, focusing on category overlap phenomena.

3.2. Basic Structure of Māori Words and Phrases

According to Biggs (1969:3), the phrase is “the natural grammatical unit” for Māori, and it has its own phrasal stress and intonation contour. A Māori phrase consists of two parts, a nucleus and a periphery. The former carries the lexical meaning and the latter the grammatical meaning. The periphery is further divided into preposed periphery and postposed periphery. Biggs also describes Māori as having two types of words: bases and particles. Bases carry lexical

7 A phrase in this context refers to what Biggs called a “contour word”: the Māori phrase that is a string of DET or TAM, base (lexical item), and the particle in the postposed periphery; it does not refer to the phrase of phrase structure rules or NP/VP in generative grammar. In this dissertation, unless otherwise specified, “phrase” used for Māori grammar means the tripartite contour word.
meanings, and therefore are used for nuclei, while particles carry grammatical meanings and are used in the peripheries. Table 3.1 illustrates the basic Māori phrase structure.

Table 3.1. Māori phrase structure

<table>
<thead>
<tr>
<th>PREPOSED PERIPHERY</th>
<th>NUCLEUS</th>
<th>POSTPOSED PERIPHERY</th>
</tr>
</thead>
<tbody>
<tr>
<td>ka TAM</td>
<td>pai ‘good’</td>
<td></td>
</tr>
<tr>
<td>te DET</td>
<td>pukapuka ‘book’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>haere ‘go’</td>
<td>mai ‘hither’</td>
</tr>
</tbody>
</table>

A contour word, or a phrase, such as ka pai ‘(is) good’, is the grammatical unit that counts for Māori grammar, not ka and pai individually. The internal structure of a phrase is strictly set. Postposed periphery slots can be filled optionally, but preposed peripheries need to be filled by either TAM or DET, with a few exceptions such as imperatives. Because the structure of the phrase is set, no grammatical rules apply phrase internally, which means that mai, for instance, will never appear in a preposed slot and TAM and DET will never follow a base (although the progressive e...ana happens to have a split form in which the second part, ana, follows the base).

The particles that occur in the preposed periphery are TAM or DET, the latter including possessive determiners such as tōku ‘my’ and demonstrative pronouns such as tēnei ‘this’. The fact that the preposed periphery has only two types of particles, TAM and DET, means that there are two types of phrases in Māori: TAM-phrases and DET-phrases.

The items that can occur in the postposed periphery are more varied. The term “particle” does not seem appropriate for some items that appear in the periphery, especially the postposed
periphery, as they can be semantically more substantial than strictly grammatical forms, even when they are difficult to translate exactly. *Rawa*, for example, is used as an intensifier:

(3.1) He roa **rawa** tēnei.

DET long too this
‘This is too long.’

*Tonu* has many meanings, one of which marks continuity:

(3.2) Kei te moe **tonu** ia

TAM sleep still 3SG
‘He is still sleeping.’

Directional particles are an important part of phrases with verbs of motion and perception:

(3.3) Haere **atu**!

go away
‘Go!’

(3.4) Kua tae **mai** te ope

TAM arrive hither DET guest
‘The travelling party has arrived.’

(3.5) Whakarongo **mai**!

listen hither
‘Listen (to me)!’

Locative “particles” also appear in the postposed peripheral position:

(3.6) **i** te **rā** **nei**

at DET day here
‘today’

In general, the items that occur in the postposed periphery seem to function as modifiers in the sense that they add more detail to the meaning carried by the base.

Although the phonological phrase and syntactic phrase do not always coincide,\(^8\) this three-slot structure is strictly observed.

---

\(^8\) For example, *Kua tae mai ia* ‘He has arrived’ can be considered to have one stress or two stresses. See Pearce (2013) for more detail.
3.3. Basic Morphosyntax of Māori

In this section, I will briefly explain the basic Māori sentence structure and the constituents used in the sentence. Māori is a VSO language.

(3.7) Kua pānui te tamaiti i te pukapuka
TAM read DET child DO DET book
‘The child has read the book.’

In (3.7), the verb pānui ‘read’ is in the sentence-initial position preceded by a Tense/Aspect/Mood marker (TAM) kua. The subject of the sentence, tamaiti ‘child’, follows the verb, and is then followed by the object pukapuka ‘book’, which is marked by the DO marker i. Both tamaiti ‘child’ and pukapuka ‘book’ are marked by the singular definite determiner te.

Thus, the sentence in (3.7) consists of three phrases: kua pānui ‘have read ’, te tamaiti ‘the child’, and i te pukapuka ‘the book (direct object)’. The nucleus, the central part of the phrase, contains the lexical meaning. Thus, pānui, tamaiti, and pukapuka are the nuclei. The periphery carries the grammatical meaning and it either precedes or follows the nucleus. The periphery of the first phrase is a TAM, kua. The second phrase has a periphery te. The last phrase, i te pukapuka, has two particles in the periphery, the direct object marking particle i and the determiner te. Māori grammar depends upon these particles in the periphery to mark the grammatical roles of the nuclei and to form phrases.

3.3.1. Determiners

As mentioned above, in Māori, lexical items rarely appear in sentences without particles such as TAM markers, determiners, and object markers:

(3.8) *Kα pānui tamariki i pukapuka
TAM read children DO book
‘Children read books.’

Example (3.8) is ungrammatical because tamariki and pukapuka are unaccompanied by
determiners. The lexical items used as arguments must be attached to some type of determiner, except in the case of object incorporation, which will be discussed in example (3.15) later in this section. The Māori determiners are shown in table 3.2.

**Table 3.2. Māori determiners**

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite</td>
<td><em>te</em></td>
<td><em>ngā</em></td>
</tr>
<tr>
<td>Indefinite</td>
<td><em>he</em></td>
<td><em>he</em></td>
</tr>
<tr>
<td>Indefinite</td>
<td><em>tētahi</em></td>
<td><em>ētahi</em></td>
</tr>
</tbody>
</table>

Note that Māori lexical items do not show a singular-plural distinction except in a handful of cases. Normally, the difference is marked by the use of the singular or plural determiner. Some examples of the exceptions to this rule, which usually refer to people, are shown in (3.9).

(3.9)  
*te tangata/ngā tāngata*  
*te wahine/ngā wahine*  
*te tamaiti/ngā tamariki*  
‘the person/the people’  
‘the woman/the women’  
‘the child/the children’

Example (3.10) shows both a plural lexical item, *tamariki*, and the plural determiner *ngā*:

(3.10)  
*Kua pānui ngā tamariki i ngā pukapuka*  
TAM read DET children DO DET book  
‘The children have read the books.’

The determiners *te* and *ngā* are usually labeled definite determiners. Although they are often translated as ‘the’ in English, it should be noted that *te* and *ngā* can be used in non-definite contexts. For example, an abstract noun can follow *te* in a general context:

(3.11)  
*He taonga tonu te wareware*  
DET property constant DET forget
‘Forgetfulness is an enduring possession.’

Indefiniteness is expressed by one of three determiners: *he*, *tētahi* (singular), and *ētahi* (plural).

*He*, referred as a “non-specific article” by Harlow (2001), has some restrictions when used in an argument. It cannot follow prepositions, including the direct object marking *i*:

(3.12) *I hoko ia i he tīkiti
   TAM buy 3SG DO DET ticket
   ‘I bought a ticket.’

Instead, *tētahi* is used:

(3.13) I hoko ia i tētahi tīkiti
       TAM buy 3SG DO DET ticket
       ‘He bought a ticket.’

Or a passive construction is used, in which *he* marks the subject:

(3.14) I hoko.na e ia he tīkiti
       DET buy.PASS by 3SG DET ticket
       ‘A ticket was bought by him.’

Another way to avoid the sequence of *i* and *he* is object incorporation, in which the direct object follows immediately after the verb without the DO marker *i* or a determiner *te/ngā*:

(3.15) a. *Ka haere ia ki te moana ki hī i he ika
       DET go 3SG DET sea to catch DO DET fish
       ‘He goes to the ocean to catch fish.’

       b. Ka haere ia ki te moana ki hī ika
       DET go 3SG DET sea to catch fish
       ‘He goes to the ocean to catch fish.’

*He* is used with the subject in intransitive sentences (3.16) and with stative verbs (3.17):

(3.16) I oma he kurī
       DET run DET dog
       ‘A dog ran.’

(3.17) I pakaru he wini
       DET broken DET window
       ‘A window is broken.’
**He** can also be used in the predicate nominal:

(3.18)  **He** tākuta ahau  
**DET** doctor **1SG**  
‘I am a doctor.’

The determiner *te* does not co-occur with personal names. However, personal names, when used as subjects, do require another marker: *a*.

(3.19)  *I* tae mai Hone  
**DET** arrive **heither** Hone

I tae mai *a* Hone  
**DET** arrive **hither** **PERS** Hone  
‘Hone arrived here.’

The personal marker *a* is not required after the equative particle *ko*:

(3.20)  **Ko** Mere taku ingoa  
**EQ** Mere my name  
‘My name is Mere.’

This particle is also used with pronouns when preceded by certain prepositions, such as *i* and *ki: i a ia* ‘to him’ and *ki a au* ‘to me’. *A* is also used with location nouns in the subject position:

(3.21)  **He** taone nui a Te Whanga-nui-ā-Tara  
**DET** city **big** **PERS** Wellington  
‘Wellington is a big city.’

However, with the names of places, *a* is not necessary after prepositions: *ki a Te Whanga-nui-ā-Tara* ‘to Wellington’.

When the lexical item refers to a location and is used in a subject position, this particle is required:

(3.22)  **He** wera a waho  
**DET** hot **PERS** outside  
‘Outside is hot.’

### 3.3.2. Prepositional particles

In example (3.7), which is repeated in (3.23), the direct object is marked by the
preposition *i*:

(3.23) Kua pānui te tamaiti *i* te pukapuka
TAM read DET child DO DET book
‘The child has read the book.’

Assuming that (3.23) is a transitive construction, *te tamaiti* ‘the child’ is the agent of the verb phrase *kua pānui*, and *te pukapuka* ‘the book’ is the direct object, marked by *i*. This particle, *i*, can also mark the agent of a neuter verb. Neuter verbs, sometimes called stative verbs, express a particular state that is the result of some action. In (3.24), *pakaru* ‘(is) broken’ describes the state of the window because of Hone’s action:

(3.24) I pakaru te wini *i* a Hone.
TAM broken DET window by PERS Hone
‘Hone broke the window.’

This particle *i* has several other usages. It serves as an oblique marker with various meanings:

(3.25) Source of motion verb
Kua tae mai te ope *i* Rotorua
TAM arrive hither DET travellers from Rotorua
‘The travellers came from Rotorua.’

(3.26) Location where the action takes place
Kei te ua *i* Rotorua
TAM rain at Rotorua
‘It’s raining at Rotorua.’

(3.27) Modifier of locative noun
Ko Rewi kei roto *i* te kāpata
EQ Rewi P in of DET closet
‘Rewi is outside the closet.’

(3.28) Time expression
I mutu te mahi *i* te rima karaka
TAM end DET work at DET five o’clock
‘The work ended at five o’clock.’

(3.29) Reason/Cause
Kua mākū te kākahu *i* te ua
TAM wet DET clothes because DET rain
‘The clothes got wet because of the rain.’
To sum up, the particle *i* can function not only to mark a direct object (3.23), but also to mark an agent (3.24), a cause (3.29), or an oblique object with various semantic roles (3.25–3.28).

Another versatile particle, *ki*, is used in similar contexts as *i*:

(3.30) Goal of motion verbs
Kei te haere ia  *ki* te taone
*TAM* go 3SG to DET town
‘He is going to town.’

(3.31) Theme object of experience verbs
Titiro *ki* te pikitia!
look at DET picture
‘Look at the picture!’

(3.32) Indirect object (recipient)
Kei te kōrero a  Hata *ki* a Rewi
*TAM* talk PERS Hata to PERS Rewi
‘Hata is talking to Rewi.’

(3.33) Location
Ka tūtaki ngā tāngata *ki* te marae
*TAM* meet DET people at DET marae
‘The people meet at the marae.’

Recall example (3.24), repeated here in (3.34), in which the particle *i* is marking the agent of the neuter verb.

(3.34) I pakaru te wini *i* a Hone.
*TAM* broken DET window by PERS Hone
‘Hone broke the window.’

The agent in the canonical transitive construction in (3.23), *te tamaiti* ‘the child’ is unmarked; that is, it appears without any prepositional particle to mark its agent role. Here is the example again, as (3.35):

(3.35) Kua pānui te tamaiti *i* te pukapuka
*TAM* read DET child DO DET book
‘The child has read the book.’

Example (3.36) is a so-called passive sentence, derived from (3.35):

"Example (3.36) is a so-called passive sentence, derived from (3.35):"
In the passive construction, the agent is marked by the particle e. To sum up, the agent can be unmarked in a canonical transitive sentence, can be marked by i in a neuter verb sentence, and must be marked by e in a passive sentence.

3.3.3. TAM

TAM (tense/aspect/mood) markers are required (except in imperatives) for the verbal phrase, whereas DET (determiners) and prepositional particles are required for items in the argument and oblique positions. Shown in the list below are some of the major TAMs:

(3.37)

(3.36) Kua pānui.tia te pukapuka e te tamaiti
        TAM read.PASS DET book       by DET child
        ‘The book has been read by the child.’

Both sentences (3.35) and (3.36) contain the TAM kua, for instance, showing that the action of pānui ‘read’, whether in active or passive form, is completed. Kua does not refer to any specific tense. It denotes that something has changed from one state to another.

(3.38) Kua mate ia
        TAM die     3SG
        ‘He has died (and is dead).’

Therefore, given the specific context, the present state of the complete action, rather than the fact of its completion, is emphasized.

(3.39) Kua oma a Rewi
        TAM run   PERS Rewi
        ‘Rewi is running (has begun to run).’

For instance, (3.39) could be uttered in a conversation between two people watching a marathon
or triathlon.

*Ka* is the most often used TAM, although it does not designate any tense, aspect, or mood. *Ka* is purely a verb phrase marker. Its interpretation depends upon other temporal markers available in the sentence:

(3.40) Ā te waru karaka **ka** hoki mai a Hone
at(FUT) DET eight o’clock TAM return hither PERS Hone
‘At eight o’clock, Hone will arrive.’

(3.41) I te waru karaka **ka** hoki mai a Hone
at(PST) DET eight o’clock TAM return hither PERS Hone
‘At eight o’clock, Hone arrived.’

It also depends upon the context:

(3.42) **Ka** timata te kura āpōpō
TAM begin DET school tomorrow
‘The school will begin tomorrow.’

According to Harlow (2001:52) TAM *i* is the only verbal particle that is “pure time”: “*I* simply locates an event or state named in the nucleus of its phrase as being in the past.”

(3.43) I haere ia ki te pikitia
DET go 3SG to DET movie
‘She went to the movie.’

Unlike *ka*, it does not require any temporal expression for the past interpretation.

*Kei te* and *i te* express the progressive aspect. The former is for present and future, while the latter is for past.

(3.44) **Kei te** uaua
TAM rain
‘It is raining.’

(3.45) **I te** uaua
TAM rain
‘It was raining.’

**E…ana**, a combination of the preposed *e* and the postposed *ana*, is similar to *ka*, in the
sense that it does not carry any tense, but e...ana is a progressive marker, used for ongoing action:

(3.46) E haere ana koe?
       TAM go TAM 2SG
   ‘Where are you going?’

It also can be used with a verb of experience:

(3.47) E mōhio ana ahau ki a ia
       TAM know TAM 1SG to PERS 3SG
   ‘I know him.’

It can co-occur with a temporal expression to denote future and past events:

(3.48) Ā tērā pō e haere ana mātou ki te moana
       at(FUT) that night TAM go TAM 1PL.INCL to DET sea
   ‘We are going to the beach tonight.’

(3.49) I tērā pō e haere ana mātou ki te moana
       at(PST) that night TAM go TAM 1PL.EXCLto DET sea
   ‘We were going to the beach tonight.’

It should be noted here that tense in Māori is not a property of verbal particles only. Because some TAMs lack tense, prepositions such as ā for future (3.48) and i for past (3.49) contribute to the tense interpretation.

Finally, lexical items cannot form verb phrases without TAM in Māori, except in imperatives and noun incorporation.

(3.50) Haere mai!
       Come hither
   ‘Come here (Welcome)!”

(3.51) Kei te kimi whare rātou
       TAM DET search house 3PL
   ‘They are house-hunting.’

We have seen that Māori verbs (the nuclei that take TAMs) do not have tense inflection. Preposed verbal particles denote tense, aspect, and mood, and a lack of tense in a TAM can be
supplemented by other temporal expressions. It also should be noted that, as seen in all the examples above, Māori verbs do not show number and person agreement.

3.4. Māori Parts of Speech

In the previous section, I presented the basic Māori simple sentence structure, which consists of the combination of the small particles that carry grammatical meaning (periphery) and the lexical items with semantic meaning (nucleus). We saw that Māori words do not have a singular/plural distinction, except for a few words referring to people. Numbers and definiteness are conveyed by determiners that precede the words (preposed periphery). Their grammatical relationships (e.g., subject, object, oblique) to the verbs are indicated by small prepositional particles that attach to the determiners and the lexical items. Verbs do not inflect for number or person, and tense, aspect, and mood are expressed by TAM, the particles that precede the verbs. Once again, it should be emphasized that the determiners and the TAM are obligatory in Māori grammar with a few exceptions (i.e., object incorporation and imperatives).

In his 1969 reference grammar of Māori, Biggs described the unit composed of a periphery and a nucleus as the basic grammatical unit:

The phrase, not the word, is the unit of Māori speech which must be emphasized in learning. It is the natural grammatical unit of the language, and even more importantly, it is the natural pause unit of speech. (Biggs 1969:3)

Harlow (2007:24) concurs and adds: “It is these particles more than the category of the lexical head of a phrase which determines the phrase’s category.”

Given that the phrase is the basic unit, along with the fact that lexical items in Māori need to be accompanied by either a TAM or a determiner, Māori has two phrasal categories: verbal
phrases (with TAM marking) and nominal phrases (with DET marking). This does not mean that
the nucleus of a nominal phrase is necessarily a noun, nor that the nucleus of a verbal phrase is
necessarily a verb. In this section, I will show some examples of such cases, which have led to
much discussion on Māori parts of speech.

3.4.1. Lexical items in nominal phrases

Māori nominal phrases are always introduced by determiners. Determiners are
undiably a property of noun phrases, and their function is to modify nouns or limit the referent
of nouns. In example (3.7), tamaiti ‘child’ and pukapuka ‘book’ both follow a determiner te.
Both lexical items are canonical nouns according to prototype analysis (Croft 1990, 2001, 2004),
as they are both used for object reference. Combined with te, they form noun phrases. The
determiners, however, can be used with items that are not canonical nouns:

(3.52) **Te** makariri o **te** wai!

DET cold of DET water

‘How cold the water is!’

*Te*, when followed by *makariri* ‘cold’—an adjective in English—expresses exclamation in this
construction. *Makariri* can also follow the indefinite *he*:

(3.53) **He** makariri te **moana**

DET cold DET sea

‘The ocean is cold.’

*Makariri* can also take a TAM:

(3.54) **Kei te** makariri ahau

TAM cold 1SG

‘I am cold.’

*Makariri* can modify a noun as well, as in *kai makariri* ‘cold food’. When these types of
utterances are taken into consideration, what Biggs and Harlow suggest sounds convincing:
Regardless of the word class of the nucleus, the particle in the preposed periphery—TAM or
DET—will determine the phrase category. It is important to note that their analysis would only claim that *makariri* in (3.52) and (3.53) appears in nominal phrases and that *makariri* in (3.54) is in a verbal phrase. It does not suggest that *makariri* in (3.52) and (3.53) is a noun and in (3.54) a verb.

Māori indeed has a great number of lexical items that can co-occur with both types of particles.

(3.55) a. *Ka pai te kōrero.*
   TAM good DET talk
   ‘The talk is good.’

(3.55) b. *I kōrero te kaiako ki ngā ākonga.*
   TAM talk DET teacher to DET student
   ‘The teacher talked to the students.’

(3.56) a. *Kua oti te kanikani*
   TAM end DET dance
   ‘The dance is over.’

(3.56) b. *I kanikai mātau*
   TAM dance 1SG.EXCL
   ‘We danced.’

(3.57) a. *Ka oti te mahi*
   TAM finished DET work
   ‘The work is over.’

(3.57) b. *Kei hea koe e mahi ana?*
   at where 2SG TAM work TAM
   ‘Where do you work?’

The pairs in examples (3.55-3.57) demonstrate that the same lexical items can be used in both nominal and verbal phrases. The English language has a similar phenomenon in pairs such as *talk* (n): *talk* (v), *dance* (n): *dance* (v), *work* (n): *work* (v). In English, this is considered zero-derivation, which is defined as a word formation process in which there is no change to the form. English, however, also possesses various suffixes for word derivation processes, such as *glory.*
(n): glorify (v), terror (n): terrorize (v), and zero-derivation is thus one of various word
derivation processes. Māori, in contrast, has an extremely limited morphology for word
formation strategies, and noun-verb conversion, termed “stem nominalization,” occurs quite
freely without any change to the forms. As the English glosses in (3.58-3.59) show, this process
appears to be equivalent to the use of English gerunds or infinitives. In (3.58), patu ‘kill’ is used
in the verb phrase, whereas (3.59) shows the result of stem nominalization.

(3.58) **Ka patu ngā tāngata i te tohorā**
TAM kill DET people DO DET whale
‘People kill the whale.’

(3.59) Kei te pīrangia a Mere kia mutu te patu tohorā.
TAM want PERS Mary TAM end DET kill whale
‘Mary wants the killing of whales to stop.’

Such nominalization is used widely, as some expressions and constructions require the
nominalized form:

(3.60) **Ka haere ahau ki te taone ki te hoko kai**
TAM go 1SG to DET town to DET buy food
‘I go to town to buy food.’

(3.61) Kua tata ngā tamariki te haere ki te kura.
TAM near DET children DET go to DET school
‘The children are just about to go to school.’ (Lit: The children have neared the going to the school)

(3.62) **He tere rawa te oma a Hata.**
DET fast too DET run of Hata
‘Hata is/was running too fast.’ (Lit: Hata’s running is/was too fast)

*Hoko kai* ‘buy food’ in (3.60) and *haere* ‘go’ in (3.61) attach to the determiner *te*, and are thus
used in the nominal phrase. In (3.62), there are two stem nominalizations involved: *he + tere* ‘a +
fast’ and *te + oma* ‘the + run’. This nominalization process allows any lexical item to be
nominalized so that it can be used in nominal phrases. In the next example, the interrogative *aha*
‘what’ is used with the indefinite determiner *he*, and the answer *whakaako* ‘teach’ in (3.63) is
also used in the nominal sense.

(3.63) **He aha tāu mahi?**
DET what your work
‘What is your work?’

(3.64) **He whakaako i te reo Māori**
DET teach DO DET language Māori
‘(It is) to teach Māori language.’

In sum, Māori determiners can be attached to items that are not nouns to form noun phrases. We saw that *makariri* ‘cold’ co-occurs with both definite and indefinite determiners, *te* and *he*. In addition, when specific constructions/expressions require a nominal phrase, any lexical item can follow a determiner through stem nominalization. It is important to emphasize again that Biggs and Harlow do not suggest that all the lexical items that co-occur with determiners are nouns, but they do maintain that phrases led by determiners are noun phrases.

### 3.4.2. Lexical items in verbal phrases

As determiners are the markers of nominal phrases, tense/aspect/mood particles are the markers of so-called verbal phrases. Examples in the previous section showed that the property-denoting noun modifier *makariri* ‘cold’ can follow determiners, and, in (3.54), repeated here as (3.65), that it can follow TAM.

(3.65) **Kei te makariri ahau**
TAM cold ISG
‘I am cold.’

Lexical items with meanings expressed by adjectives in English are used either in nominal phrases, as seen in the previous section, or in verbal phrases in Māori:

(3.66) **Ka tika**
TAM right
‘(That)’s right.’
Kua riri te tupuna
TAM angry DET grandfather
‘Grandfather is angry.’

Kei te hiakai ngā tamariki tāne
TAM hungry DET boys
‘The boys are hungry.’

Kei te pēhea koe?
TAM how 2SG
‘How are you?’

Kei te pai
TAM good
‘(I’m) good.’

Note that the interrogative pēhea ‘how’ is used in the verbal phrase in (3.69).

The verb phrases appear to be more rigid than the noun phrases in terms of the choice of nucleus. That is, not all lexical items can follow TAM to be used in verbal phrases. Lexical items that denote objects, people, and places are not readily used with TAM in verb phrases. Biggs (1969) considers that the lexical items that do not appear in verbal phrases belong to the class of nouns, and his examples include words such as ika ‘fish’, kiore ‘mouse’, rākau ‘tree’, because, according to Biggs, ka rākau, for instance, is unheard of. However, this is argued against by Muturangi (1979:31), whose grammatical judgment permits utterances such as:

Ka rākau katoa ngā taha o te maunga ra
DET tree all DET side of DET mountain there
‘The slopes of that hill are all covered with trees.’

E ngaru ana te moana i te rā nei
TAM wave TAM DET ocean at DET day this
‘The ocean is rough today.’

Kua tangata te tamaiti
TAM man DET child
‘The child has become a man.’ (Hohepa 1967:49)

The use of objects such as rākau ‘tree’ and ngaru ‘wave’ in verbal phrases has been accepted by
Harlow:

(3.74) I kōhatu.ngia a Pānia
DET stone.PASS PERS Pānia
‘Pānia turned into stone.’ (Harlow 2001:188)

(3.75) Kua tangatawhenua-tia te ope
TAM natives.PASS DET group
‘The visiting party has become the natives.’ (Harlow 2001:188)

(3.76) Kua whare.tia te taone⁹
TAM house.PASS DET town
‘The town is covered with the houses.’

Notice that, in the three examples above, the passive suffixes -ngia and -tia are used in the verbal phrases and the past tense or perfective are used for TAM. The informant who gave (3.76) accepted this sentence only with the perfective kua and with the passive, while Harlow accepts the past tense marker i. Muturangi (1979), on the other hand, accepts the use of object-denoting words in verbal phrases with the TAM non-tense, non-aspect marker ka as well as with non-tense progressive e...ana, and without a passive suffix attached. Grammaticality judgments can be affected by the dialect, fluency, and socio-economic background of the speakers. Moreover, the use of nouns in verbal phrases is different from the use of verbs in nominal phrases. Clark and Clark (1979) claim that the use of nouns as verbs is regulated by convention. When a speaker uses a noun as a verb, as in “I penciled you in for next Friday,” or “I will xerox this,” or “Do you want to super-size it?” the speaker’s intention is understood and accepted by the listeners who share the same knowledge. The speech participants’ shared knowledge allows the meaning of the parent noun to be successfully transferred to the verbal use. Unlike stem nominalization, which is a syntactic process that can apply to almost any lexical item regardless of the meaning and that

⁹ This usage is accepted also by Dr.Wiri.
does not change meaning in the nominalized word, the process by which nouns are used as verbs is a derivational process that has certain semantic effects. For denominal verbs to be created, mutual understanding of the meaning by the speech community must be established. For example, some animal nouns like *dog* and *fox* can be used as verbs, as in “I dogged her” (track or follow like a dog) and “She foxed me into doing this” (deceive like a fox), and the speech participants understand the connotations based on their shared knowledge. As such derivations are not available for all animal nouns, the verbal usage of nouns is an innovative process. The less frequent use of nouns in verbal phrases is attributed to these characteristics of denominal verbs.

Derived verbs in English enter the inflectional paradigm: *cow, cows, cowed, cowing*. Thus, the word forms make it evident that a derivational process took place. In Māori, however, the lack of inflectional morphemes leaves only the particles to show their grammatical roles. The denominal verbs are marked with the available verbal morphemes, such as affixes and TAM markers, to convey the meaning. In examples (3.74) to (3.76), the meaning is that “something (the original noun) changed into or became something else.” The passive suffixes\(^{10}\) -*ngia* and -*tia* are used with the past and perfective TAMs to bring out the meaning of completed action and its lingering effects. Muturangi (1979) and the speech community to which he belongs share a different knowledge that accepts the usages seen in examples (3.71) and (3.72). The characteristics of the denominal process (i.e., it is an innovative process that is accepted by the speech community, rather than a syntactically triggered process) can be seen in the varied judgments of the speakers.

\(^{10}\) The passive suffix -*Cia* is discussed in the section 3.5.
3.4.3. Parts of speech or not?

When one faces data in which verbs do not have to be the head of the nucleus of verbal phrases and nouns do not have to be the head of the nucleus of nominal phrases, the account proposed by Biggs and Harlow becomes convincing. To recap their analysis, in Māori, it is the particles, not the category of lexical items, that assign the phrasal categories, and the phrase, not the word, is the basic grammatical unit. The fact that lexical items such as *pai* ‘good’ and *tamaiti* ‘child’ are not used in isolation in sentences is strong evidence that the particles to which they attach define their categories. The fact that many lexical items can be used freely in both nominal and verbal phrases strongly suggests that the categories of the lexical items are not important, at least syntactically. These facts support Biggs’s and Harlow’s analyses. However, Māori has productive affixes that select specific phrase markers, DET or TAM. In the next section, I will discuss two Māori suffixes that imply categorial distinction on the lexical level in Māori.

3.5. -Canga and -Cia

Although Māori does not have inflectional morphemes, there are some highly productive derivational affixes, such as -Canga and -Cia.\(^{11}\) -Canga derived words select DET, and -Cia derived words select TAM (and can be nominalized, allowing them to select DET), which suggests that lexical categories do exist in Māori.

3.5.1. Nominalizing suffix -Canga

The nominalization suffix -Canga is highly productive. According to Harlow (2007:121), -Canga “derive[s] nouns.” Bauer (1993:512) claims that -Canga can be used as “a criterion for

---

\(^{11}\) “C” in -Canga and -Cia represents a “thematic consonant.” The word-final consonants were lost in many Oceanic languages. However, the lost word-final consonant is retained in the suffixed form. For example, *ina.mia* (drink.pass) ‘be drunk’ is originally *POC (Proto Oceanic) inum-ia.*
verbhood,” meaning that the lexical items that take -Canga are verbs. This view has some counterexamples, which I will introduce later in this section.

-Canga nominalization is often seen in past time adverbial clauses:

(3.77) I tōna tae.nga mai, ka kai tātou
TAM 3SG arrive.NMLZ hither, TAM eat 1PL.INCL
‘When he arrived, we ate.’

It is also used to express cause:

(3.78) Nā te patu.nga a Hone i te tohorā, ka riri āna tamariki
P DET kill.NMLZ PERS Hone DO DET whale TAM angry his children
‘Since Hone killed the whale, his children are mad.’

Comparative expressions use -Canga nominalization:

(3.79) I tere ake tāna oma.nga i tāu
TAM fast up 3SG run.NMLZ than 2SG
‘He ran faster than you.’ (Bauer 1997:520)

-Canga forms can be subjects:

(3.80) E kore e ngaro tō puta.nga
TAM NEG TAM lost 2SG appear.NMLZ
‘Your appearance is not going to be forgotten.’

-Canga nominalized words can denote the place where an action occurs:

(3.81) moe: moe.nga ‘sleep: bed’

They can also mean the product of an action:

(3.82) tuhi: tuhi.nga ‘write: writing’

As mentioned earlier, despite Bauer’s suggestion that -Canga can establish belonging in the verb category, there are instances in which -Canga attaches to lexical items that are not characteristically verbs:

(3.83) rangatira: rangatira.tanga ‘chief: chiefliness’
Māori: Māori.tanga ‘Māori: Māori-ness’

In (3.83), the -Canga derived nouns denote the quality that the original nouns possess. It is also
well known that the modifier of a word with -Canga must also bear a variant of -Canga, which appears to be a kind of agreement:

(3.84) tō rātou tūtaki.kanga tuahi.tanga
      3PL.POSS meet.NMLZ  first.NMLZ
      ‘their first meeting’

In all of the above examples, the -Canga derived words must follow a determiner, te, or a personal pronoun such as tāna. The fact that -Canga derived words cannot appear with TAM strongly suggests that there is a category distinction at the lexical level. However, note that the existence of the category at the lexical level implied by the suffixation of -Canga does not necessarily support Bauer’s proposal for establishing a category of verbs in Māori. While she suggests that the stems that -Canga can attach to are verbs (which is problematic, as examples [3.83] and [3.84] demonstrate), the fact that -Canga derived words can only select one specific category (DET) does imply that there is a lexical category in Māori.

3.5.2. Passive suffix -Cia

The suffix -Cia is regarded as a passive suffix by most Māori grammarians:

(3.85) a. Active
      E here ana ngā tamariki i te hōiho
      TAM tie  TAM DET children DO DET horse
      ‘The children are tying the horse.’

(3.85) b. Passive
      E here.a ana te hōiho e ngā tamariki
      TAM tie.PASS TAM DET horse  by DET children
      ‘The horse is being tied by the children.’

Note that the agent subject ngā tamariki ‘children’ in (3.85) a becomes the oblique in (3.85) b marked by e ‘by’, and the direct object te hōiho ‘horse’ in (3.85) a is the upgraded subject in (3.85) b through passivization. Although the passive in Māori appear to be much like that of English in the examples in (3.85), Māori uses passives more often than English. Bauer (1993)
notes that native speakers reject the active construction if the sentence in (3.85) is in the past tense:

(3.86) *I here ngā tamariki i te hōiho
TAM tie DET children DO DET horse
‘The children tied the horse.’

Passive forms are required for imperative expressions:

(3.87) Inu.mia te miraka
drink.PASS DET milk
‘Drink milk!’

Intransitive verbs can be without passive endings:

(3.88) E oma!
‘Run!’

However, intransitive verbs can be used in passive forms as well:

(3.89) Kua oma.kia tōku oma.nga
TAM run.PASS my run.NMLZ
‘I have finished my course.’ (Lit: ‘My running was run.’) (Harlow 2007:106)

As discussed for example (3.76), nouns can be used with the passive suffix in a particular context:

(3.90) Kua whare.tia te taone
TAM house.PASS DET town
‘The town is covered with the houses.’

Harlow (2007) mentions that even proper nouns can be passivized:

(3.91) Koinei i Ngā Pōtiki.hia ai aua whenua
this TAM Ngā Pōtiki.PASS ANAPH those land
‘This is why those lands become Ngā Pōtiki.’

The passive suffix is also used with the interrogative:

(3.92) I aha.tia tō waewae?
TAM what.PASS your leg
‘What happened to your leg?’ (Lit: ‘Your leg was what-ed?’)

-Cia, like -Canga, requires modifier agreement:
In all of the examples above, the passive suffix -Cia attaches to bases that are not verbs, which is puzzling, as “passivization” is a property of verbs. However, all the passivized words appear with a TAM marker in the examples above, which suggests that the passive suffix -Cia creates the bases that belong to a single lexical category, much as the nominalizing suffix -Canga seems to define a lexical category because a lexical item with -Canga can only select DET. For the bases to which the passive suffix -Cia seems to freely attach, Biggs (1961, 1969) suggests that the word’s ability to take -Cia is the determining factor for the word class (“universal”), as -Cia does not appear with neuter verbs, such as pakaru ‘broken’ or pau ‘is used up’. Such a constraint does imply a category distinction of the base to which the suffix can attach. However, there is also a possibility that the constraint is strictly semantic.

3.5.3. Interim summary

Māori has two phrase markers, TAM and DET, which both can be followed by base items that would fit into various categories in the classic part-of-speech system. The Māori bases cannot appear in isolation in the phrases. Māori linguists have claimed that the basic grammatical category in Māori is phrasal, not the individual word. Putting all these facts together, it appears that Māori bases are not preassigned categories; however, the suffixes -Canga for nominalization and -Cia for passive suggest the possibility that Māori has categories on the lexical level, as they are not flexible in the choice of grammatical particles that can occur with the bases to which they attach. In the next section, I will introduce how the three major Māori linguists who have published Māori reference grammars approach this issue.
3.6. Proposed Word Classification Systems for Māori

In this section, the word classification systems proposed by the three main Māori grammarians, Bruce Biggs, Winifred Bauer, and Ray Harlow, will be introduced.

3.6.1. Biggs’s analysis

Biggs (1961) proposed in his dissertation that the basic grammatical unit in Māori is not the word but the phrase, which he calls the “contour word.” Māori words are divided into bases (carrying the lexical meaning) and particles (grammatical words). The nucleus contains a base that carries the lexical meaning, which can combine with a TAM or DET particle to form a verbal or nominal phrase. For Biggs, word classes in Māori should be based on the compatibility between bases and particles to avoid overlap problems. Biggs therefore offered a new classifying system based on distribution, instead of relying on the traditional part-of-speech classification.

He divided the bases (words that carry lexical meanings) into five groups:

- Universal: the bases that take passive -Cia
- Stative: the bases that can be used verbally but cannot co-occur with passive -Cia
- Locative: the bases that can occur directly following the locative particle ki without a determiner
- Personal: the bases that need the personal particle a
- Nouns: the bases that do not occur with TAM, do occur with the determiners te/nga, do not occur with -Cia or the personal particle a, and cannot immediately follow ki

According to this classification, a large number of Māori bases are universals. They can be used in both nominal and verbal phrases. For instance, tangi ‘weep’ is a universal, as it has a passive form tangi.hia ‘being wept’. Therefore, this class contains words that are nouns such as kōhatu ‘stone’, for it can take the passive ending, as in kōhatu.tia ‘be stoned (become a stone)’.
Moreover, although Biggs considers *rākau* ‘tree’ as a noun because it lacks a passivized form, there are examples in which it is passived and used with a similar meaning as in (3.90).

(3.94) Kau *rākau.tia* te taone
     TAM tree.PASS DET town
     ‘The town is covered by the trees.’

Based on its use in this example, *rākau* would be a universal, and words such as *aha* ‘what’ and *pēhea* ‘how’ are also universals because of their ability to co-occur with -Cia.

The examples Biggs gives for the stative category are *ora* ‘well’ and *tika* ‘correct’. He explains that statives express a state, condition, or attribute (1969:19). They can be used in both nominal and verbal phrases, but unlike the universals, the statives cannot take the passive -Cia suffix. As Biggs eliminated the category of adjectives, some of the adjectives are then categorized as statives:

(3.95) Ka *tika* te kōrero
     TAM correct DET talk
     ‘The talk is correct.’

Some statives have passive meanings, as seen in the next example from Biggs:

(3.96) Ka *wera* te whare i te ahi
     TAM destroyed DET house by DET fire
     ‘The house was destroyed by fire.’ (Biggs 1969:54)

Note that *wera* means ‘destroyed’ without a passive ending, and the agent is marked by *i* rather than *e*, the agentive marker in the passive construction.

Biggs defines locatives as bases that refer to position in space and time and that never take definite or indefinite articles when they follow locative particles such as *ki, kei, i, and hei.*

(3.97) Kei *roto ia i te marae*
     TAM inside 3SG at DET marae
     ‘He is in the marae.’
(3.98) Haere mai ki konei
come hither to here
‘Come here!’

(3.99) I haere a Hone ki tātahi
TAM go PERS Hone to beach
‘Hone went to the beach.’

(3.100) Kei Pōneke ia e noho ana
TAM Wellington 3SG TAM stay TAM
‘He lives in Wellington.’

Biggs posits a separate category, personals, for bases that require the personal article a after the locative ki:

(3.101) Ka kōrero ia ki a Pita
TAM talk 3SG to PERS Pita
‘He/she talks to Pita.’

He separates the two classes of locatives and personals from the nouns, which, as mentioned earlier, leaves only a very limited number of nouns. Because his classification relies on one major determining factor—the bases’ compatibility with the passive -Cia—many bases are categorized as universals. However, as shown in (3.91), some locatives (and personals) can appear with -Cia.

Biggs created a new classification system that fits Māori by using only distributional criteria and disregarding the semantic aspect of the words. The primary factor for word classification is compatibility with TAM: those that cannot co-occur with TAM are nouns. The next step is to differentiate the items that do take TAM but show different behaviors, such as the ability to take -Cia. The result is that the newly created category of universals contains a large number of lexical items, which include nouns, verbs, and adjectives. Biggs’s analysis of the phrase as a contour word has influenced other Māori linguists. However, other linguists have found his proposal of a large group of words that includes nouns, verbs, and adjectives and is
defined by a single factor—compatibility with one suffix—to be problematic.

3.6.2. Bauer’s analysis

Bauer (1993, 1997) maintains Biggs’s ideas regarding phrases as the most important grammatical unit in Māori and the existence of two major word classes, particles and lexical bases. Bauer differs from Biggs in her solution to the problem of Māori words’ overlapping nature by “accepting that most base forms of words in Māori have the potential to be used in either nominal or verbal constituents, with a corresponding difference in sense” (Bauer 1997:65). She points out that some syntactic phenomena in Māori suggest the traditional part-of-speech distinctions, and she suggests that the overlap of word classes is the result of zero-conversion.

Bauer proposes the following word classes for Māori:

- Nouns
- Pronouns
- Verbs
- (Adjectives)
- Prepositions
- Numerals/Quantifiers
- Personal nouns
- Locative nouns
- Verbal particles
- Definitives

Bauer finds Biggs’s criteria for nouns (i.e., they co-occur with the determiners, but not with TAMs, the passive suffix, or the personal particle a) too restrictive, while she appears to agree with Biggs in excluding proper names and location nouns from the category of nouns.
Whether they are in separate word classes or are subclasses of nouns is inconclusive in her discussion. Unlike Biggs, Bauer posits a category of verbs, although she does admit that “there do not appear to be clear-cut structural properties which could help to provide an operational definition for the class verb, because of the functional overlap with non-verbal predicates” (Bauer 1993:259). She disagrees with Biggs’s “universal” category, which includes all the bases that can occur with TAM and a passive suffix, and also claims that Biggs’s stative category based on incompatibility with the passive suffix -Cia does not work because ora ‘well’, for example, can take -Cia. She suggests that Biggs’s statives, which are also known as neuter verbs, should be a subcategory of the verbs. Therefore, although she arrives at no clear definition for verbs, she assumes the category exists.

Biggs’s word classification system does not include adjectives. Bauer notes that “it is doubtful whether there is any operational definition of the adjective phrase, and possibly doubtful whether this category is found in Māori” (Bauer 1993:95). She offers some constructions that can identify adjectives. For instance, only adjectives can be modified by tino ‘very’, as in tētahi whare tino nui ‘a very big house’, but *tētahi whare tino miraka ‘a very milking shed’. She also suggests that adjectives cannot be strung together, as shown in her example:

(3.102)tētahi tangata tino nui, tangata tino mōmona
  one man very big man very fat
  ‘a very big and fat man’ (Lit: a man very big, man very fat) (Bauer 1993:107)

She further suggests that only adjectives can appear in comparative and superlative structures:

(3.103)He roa atu ōna makawe i ōku
  DET long away his hair than my
  ‘His hair is longer than mine.’

She considers these criteria as a good “potential operational definition” (Bauer 1993:260) for
adjectives, but also warns her readers that the words her criteria identify as adjectives are not always the same as those identified by other linguists. Therefore, she suggests that only those words identified by all the criteria that linguists have suggested be considered as forming the group of adjectives. In the end, however, she concludes that adjectives, as well as statives, should be a subclass of verbs, because they can co-occur with TAM. Hence, in sum, Bauer appears to concur with Biggs that adjectives do not form their own category.

3.6.3. Harlow’s analysis

Harlow (2001, 2007) identifies Māori word classes by “grouping together those items which share privileges of occurrence and/or can undergo the same morphological processes” (2007:96), due to the paucity of inflectional morphology and difficulty of word classification. He takes the same position as Bauer that the overlapping of the word categories is the result of zero-conversion. Harlow agrees with Waite (1994), who claims that Māori functional categories, I(nflection) and D(eterminer), can co-occur with NP, VP, and AP as their complement.

According to this interpretation, it is not that pai ‘good’ is assigned to multiple parts of speech and becomes a noun when it appears with DET or a verb when it appears with TAM. Rather, the lexical item has a preassigned category (which is determined by speaker’s instinct, according to Waite), and undergoes zero-derivation when necessary. Harlow claims that there are ways to identify the categories based on words’ syntactic and morphological behavior.

Harlow follows Biggs in claiming that Māori has verbal phrases and nominal phrases. He further divides the nominal phrases into two subclasses: one led by determiners and the other led by prepositions. He also follows Biggs in dividing Māori words into two large groups, bases and particles, while he agrees with Bauer in restoring nouns and verbs as categories for the bases. Harlow divides the nouns into three subcategories and the verbs into five.
The subclasses of nouns are common nouns, locative nouns, and personal nouns. Pronouns are in the subcategory of personal nouns. Harlow notes that “the common nouns *par excellence* are those derived by means of the nominalizing suffix” (Harlow 2007:110), and mentions that a “large number of other items, both simple and compound, belong to this class.” He does not give much further definition for common nouns, except that they are introduced by the determiner *ngā*, while the locatives and personals are identified by the same criteria suggested by Biggs and Bauer. Biggs give separate categories for the locatives and personals, while Bauer and Harlow consider them as subclasses of nouns.

The verbs are further divided into transitive, intransitive, middle, neuter, and adjective, all of which can co-occur with TAM. Harlow bases this division on argument structures, syntactic operations that the verbs can undergo and the strategies involved (such as passivization and relativization), semantic aspects (such as verbs of motion, verbs of perception, etc.), and compatibility with other particles. He offers a summary of the different behaviors of the five verbal subclasses, shown in table 3.3 (Harlow 2007:108):
Table 3.3. Classes of verbs

<table>
<thead>
<tr>
<th></th>
<th>Number of arguments</th>
<th>a-/o-</th>
<th>Imperative</th>
<th>Can modify</th>
<th>Passive suffix</th>
<th>Actor emphatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transitive</td>
<td>2</td>
<td>a-</td>
<td>e/∅</td>
<td>Y</td>
<td>Y</td>
<td>y</td>
</tr>
<tr>
<td>Middle</td>
<td>2</td>
<td>o-</td>
<td>kia</td>
<td>Y</td>
<td>Y</td>
<td>n</td>
</tr>
<tr>
<td>Intransitive</td>
<td>1</td>
<td>a-/o-</td>
<td>e/∅</td>
<td>Y</td>
<td>Y</td>
<td>y/n</td>
</tr>
<tr>
<td>Neuter</td>
<td>1</td>
<td>o-</td>
<td>kia</td>
<td>N</td>
<td>N</td>
<td>n</td>
</tr>
<tr>
<td>Adjective</td>
<td>1</td>
<td>o-</td>
<td>kia</td>
<td>Y</td>
<td>N</td>
<td>n</td>
</tr>
</tbody>
</table>

Harlow relies on careful observation of (mostly) syntactic behaviors to divide the verbs into subclasses. However, what holds all five subclasses together in one category is simply the ability to co-occur with TAM.

3.7. Parts of Speech or Not?

As discussed in section 3.6, the Māori linguists all agree that words in Māori are hard to classify according to the traditional parts-of-speech system. They all concur on the following points:

1. The difficulty of word classification in Māori is due to the lack of inflectional morphology.

2. The basic grammatical unit in Māori is a phrase, not an individual lexical item.

---

12 Māori has two possessive prepositions, a and o, as in te wahine a Hone ‘Hone’s wife’ or te whare o te tangata ‘the man’s house’. The selection between a and o depends upon the relationship between the possessor and possessum (the thing possessed). Simply put, if the possessor has control over the possessum, a is selected. This a/o distinction appears to mark the subjects in nominalized clauses. When a transitive verb is nominalized, a is used to mark the subject: te patu.nga a Hone i te tohorā (DET hit.nmlz of Hone DO DET whale) ‘Hone’s killing of the whale’. For middle and neuter verbs, o is used: te mōhio.tanga o te Mere ki te waiata (DET know.nmlz of DET Mere to DET song) ‘Mary’s knowledge of the song’.
3. All Māori phrases are either nominal or verbal, the former headed by DET, the latter by TAM.

4. Māori words are split into two large groups: particles, which are the grammatical/functional items such as DET and TAM, and bases, which are the items that carry lexical meanings and follow either DET or TAM.

Biggs established a new classification system that relies as little as possible on the traditional part-of-speech system, while Bauer and Harlow embrace the part-of-speech system and modify it in order to apply it to Māori. Although these linguists’ solutions to word classification may differ, they all discuss the categories of the bases, in spite of the fact that they also agree that the basic grammatical unit is a phrase.

Therefore, simply put, the puzzling issue is the category mismatch between the functional category (the particle) and the lexical item (the base), or the phrasal category (TAM or DET) and the lexical category (noun, verb, etc.), as seen in combinations such as *te oma* DET + ‘run’ or *e ngaru ana* TAM + ‘wave’ + TAM. Generally accepted is that a noun is the head of a phrase, with the determiner in the specifier position to form NP, or that DET takes an NP whose head is a noun as a complement to form DP. A verb is the head of VP, which is a complement of IP. The category mismatch between the particle and base, or between the phrasal category and the lexical category of the head, can be accounted for by two possible explanations: First, the category matching mechanism has a flaw, or second, the categorization itself has a flaw. Taking the latter view, Biggs (1961, 1969) proposed a Māori-specific word classification system, while Harlow (2001, 2007) and Bauer (1993, 1997) maintained the traditional part-of-speech system by making the necessary adjustments for Māori. Hengeveld (1992) claims that there are languages that do not make part-of-speech distinctions and instead possess one lumped category. In this vein, Waite (1994) suggested that Māori functional categories (TAM and DET) can take NP, VP, and
Yet there is also another possibility. As mentioned in chapter 2, some typological studies show that category flexibility in one grammatical level needs to be counterbalanced at some other level. That is, as long as the phrasal categories, which are the “basic” grammatical units, are available for the sentence structure, it may suffice for Māori grammar. In other words, if there are no lexical categories, a category-matching requirement does not exist. Languages such as Māori that are said to have flexible word class have captured linguists’ attention because the conventional approach does not apply. In the next chapter, I will examine the alternative interpretation of flexible word class: category underspecification. What benefit or edge does the part-of-speech system give us over the precategorial approach to account for the fluid word class phenomenon in Māori? This question needs to be kept in mind when one searches for an account for flexible word class.
Chapter 4

Category Undespecification

4.1. Introduction

In chapter 2, I discussed the origin and the development of the part-of-speech system and showed how loosely the concept and terminology of parts of speech have been used in the study of languages: to refer to syntactic categories and lexical categories; taken as language universals postulated a priori or language-specific word categorization systems; defined morphosyntactically or semantically; considered to be sharply demarcated categories or prototype categories with fuzzy boundaries. Typological studies show that languages have different types of part-of-speech systems, among which are systems with flexible categories and no clear distinction between the classic parts of speech. In chapter 3, I provided examples from Māori, which is known to have flexible word category, and illustrated how Māori grammar writers have dealt with the fluid category phenomenon in their accounts. Biggs (1961, 1969) created a language-specific word classification system for Māori, making a conscious effort not to refer to the traditional part-of-speech system, while Bauer and Harlow (2001, 2007) adopted the traditional part-of-speech system, assuming zero-derivation to explain category overlapping. However, one issue that Māori linguists have all agreed on is that the basic grammatical unit in Māori is a phrase, not a lexical item. This insight motivates an inquiry into the approach that assumes the absence of category on the lexical level. In this chapter, I will examine different claims regarding category underspecification. In 4.2, I will introduce a theory that does not posit inherent categorial features in generative grammar. In 4.3 to 4.8, I will illustrate how precategorial approaches have been applied to various languages. The chapter concludes with a
summary in 4.9.

4.2. Morphology without Categories

Category underspecification tends to be associated with less researched, non-Indo-European languages, but the idea already existed in generative grammar. Distributed morphology (DM), proposed by Marantz (1998) and Halle and Marantz (1993, 1994), is a theory that counters the lexicalist position that assumes the existence of a lexicon. The lexicon, as part of the base component in Chomsky’s syntactic theory, is “simply an unordered list of all lexical formatives” (Chomsky 1965:84). The lexicon provides the items for the syntactic operations of sentence derivation, and the lexicalist claim is that words are created in the lexicon, ready for syntactic derivation, in a process that is separate from the domain of syntax. However, Chomsky, in his 1970 paper “Remarks on Nominalization,” suggested

as a tentative hypothesis, that a great many items appear in the lexicon with fixed selectional and strict subcategorization features, but with a choice as to the features associated with the lexical categories noun, verb, adjective. The lexical entry may specify that semantic features are in part dependent on the choice of one or another of these categorial features. (1970:190)

DM took this stance and developed a morphological theory in which the same generative rules work for phrase formation and word formation. Thus, instead of assuming words to be formed in the lexicon, which is a part of the base component, so that they can be inserted into the derivational process, the lexical insertion occurs at different stages of the derivation, governed by syntax. DM sees no border between syntax and morphology, but syntactic hierarchical structure
all the way down. Word formation can be described in the same manner as phrase formation, using a tree structure with a binary system. However, unlike the other generative theories in which the syntactic category is assigned to each lexical item in its feature specification, DM does not assume the categories associated with the lexical items. It claims that the underlying neutral forms, called “roots,” are category-less. When the root is used in the nominal context, that is, when the nearest c-commanding category-defining morpheme is a determiner, the root appears with it as a noun. On the other hand, if the root is inserted into a verbal environment, which means the nearest c-commanding category-defining morpheme is “little” v tense and aspect, it is realized as a verb. Therefore, because there is no lexicon in the grammatical model in DM, what is inserted in the derivation is a root, which has no category specification, instead of a lexical item. For example, in *John destroyed the city*, a root √DESTROY is inserted in the structure instead of a lexical item *destroy* [V], as in figure (4.1). As the root is in the verbal configuration, it is morphologically realized as the verb (in a verbal form) *destroy* in contrast to figure (4.2), where it is realized as *destruction* in a nominal form.

![Diagram](image)

**Figure 4.1. Root √DESTORY in verbal configuration in John destroyed the city**
While DM does not question the existence of the classic part-of-speech system, it differs from other generative models in that it does not assume a one-to-one relationship between category and lexical item.

Lieber (2004, 2006) agreed with DM’s category-underspecification theories, but she was not satisfied with its unclear affixal selection processes. She came up with the lexical semantic model, in which she claims that what roots and affixes lack is syntactic category, not semantic category. Therefore, in her model, lexical items contain the information that is relevant for syntax (the semantic skeleton) and also the information concerned with the interpretation of the lexical items (the semantic-pragmatic body). The skeleton has two conceptual categories: SUBSTANCES/THINGS/ESSENCES, which is represented by the semantic feature [+/- material], and SITUATIONS, represented by [+/- dynamic]. The skeleton also contains subcategorization information (i.e., how many arguments it takes), represented by square brackets [ ]. The following is the skeleton for write:

![Diagram of the skeleton for write]
Write subcategorizes two arguments with thematic roles of agent and patient, in that order, represented by the two sets of brackets.

Affixes also have a skeletal representation. The suffix -er forms concrete processual nouns, and has an “R” argument in its skeletal representation, represented by one bracket:

(4.2) -er  [+material, dynamic ([ ], <base>)] (Lieber 2004:39)

Thus, when the two are linked in writer, the skeletal representation is as in (4.3).

(4.3) writer

   [+material, dynamic ([i], [+dynamic ([i], [ ])])) (Lieber 2004:68)

   -er                                        write

In this model, the affix has hierarchically higher status, seeking matching features and arguments to coindex. The coindexed arguments will be those relevant for syntax. For writer, the affix’s argument is coindexed with the first and closest of write’s two arguments: the one with the thematic role of agent. This process gives the derived lexical item writer the interpretation of agent. Although the suffix -er is known to be able to attach to N (e.g., village), V (e.g., write), and A (e.g., foreign), Lieber shows that the derivational process can be accounted for without referring to the syntactic categories N, V, and A. Lieber’s goal is to show the affixal mechanism, not to prove category underspecification for the derivatives. However, her account of the affixation process using semantic features demonstrates that the classic part-of-speech notations are not necessary for roots, which implies that the categories required by syntax are not inherent in the nature of the individual lexical items. As discussed in chapter 2, Baker (2003) claims that

13 “R” is the external argument of a noun, and it means “referential” (Lieber 2004:16).
categories are determined by syntax, rejecting inherent lexical categories. This insight also indicates that, whether categories are imposed by structure or defined by semantics, they become relevant at particular levels in grammar. In the next section, I will introduce the Philippine languages as an example of languages that lack noun and verb differentiation at the lexical level.

4.3. Foley (Philippine Languages)

Foley (1998) argues that Philippine languages do not have lexically specified word classes. His insight stems from the so-called focus system found in Tagalog (and other Philippine languages). In the Tagalog focus system, verbal morphology correlates with the semantic role of the syntactically prominent noun phrases, which Foley named “pivots.” The syntactically prominent noun phrase is morphologically marked with a prenominal marker *ang*. The semantically prominent noun phrase can have any semantic role—agent, patient, location, instrument, and so forth—which allows arguments that could not be subcategorized by verbs according to the standard grammatical theories to be chosen for the pivot. In English, for instance, the NP that can serve as a pivot has to be subcategorized by a verb. In the sentence, *A kid stole candy in the store*, the verb *stole* subcategorizes *a kid* as the actor and *candy* as the undergoer. In this sentence, *a kid* is a pivot, but in the passivized sentence, *candy* is the pivot: *Candy was stolen by a kid in the store*. However, in English, *in the store*, which is not a subcategorized argument, cannot serve as a pivot. In contrast, that is what is possible in Philippine languages,\(^\text{14}\) by using the verbal affixes that correspond with the NP marker to mark the pivot. Tagalog verbs are never used “unmarked” (in the sense that English *give* is unmarked);

\(^\text{14}\) For detailed discussion and examples, see Foley 1998.
they require overt verbal affixes that serve to signal which NP is the pivot in the sentence. Foley argues that Tagalog verbs in root form (i.e., unmarked, without affixes) do not carry argument structure information like English verbs do. That is, argument structures become available only when the roots are derived and the affixes are attached. Consequently, this suggests to Foley that the roots, having no argument subcategorization information, are not to be considered as lexical verbs. He claims that Tagalog roots are precategorial; they are neither nouns nor verbs. He further argues that, in spite of the superficial differences,

all words in all languages have a semantic representation of lexical conceptual structure...let us assume these are roughly similar across languages, so that English *give*, Tagalog *bigay*...all would have a lexical conceptual structure roughly along the lines of “X causes Y to have Z.” (Foley 1998:47)

Therefore, what Foley denies is the existence of a noun–verb distinction in Tagalog on the lexical level. The lack of word class differentiation also can be witnessed diachronically, as Bisang (2010, 2013) shows in the case of Late Archaic Chinese, discussed next.

4.4. Bisang (Late Archaic Chinese)

Late Archaic Chinese is the language used in the classical texts written between the fifth and third centuries BC. Bisang (2010, 2013) argues that lexical items in Late Archaic Chinese are precategorial, and provides a diachronic reason: The morphological system of word class distinction that existed in earlier Chinese disappeared in Late Archaic Chinese, leaving only the syntactic slots for the category distinction of N and V.

Bisang (2010, 2013) examines the argument structures of intransitive and transitive
verbs. In Late Archaic Chinese, the structures for the intransitive and transitive constructions are fixed, as shown in (4.4), where the subscripts S, A, and U refer to Subject, Agent, and Undergoer respectively (Bisang 2013: 279).

(4.4) Intransitive \[ \text{V NP}_S \text{ or NP}_S \text{ V} \]

Transitive \[ \text{NP}_A \text{ V} \text{ NP}_U \]

In the transitive structure, the word order is fixed as shown in (4.5), but if there is no overt agent NP, the undergoer NP can be either pre or post V. The choice of category of the lexical items that can appear in V position is flexible. The following is the intransitive structure, in which the person-denoting lexeme appears both in the subject and V positions:

(4.5) 君 君 臣 臣 父 父 子 子
jūn jūn chén chén fū fū zǐ zǐ
‘Let the prince behave like a prince, the minister like a minister, the father like a father, and the son like a son.’ (Bisang 2013:281)

In the transitive construction, any action-denoting or property-denoting lexeme with one argument position can be used, which brings the causative and putative meanings. In the next example, as Bisang explains, the property-denoting verbs 小 xiăo ‘small’ and 大 dà ‘big’ (underlined) are in a transitive construction and give the causative ‘make it small/big’ interpretation.

(4.6) 鼻 大 可 小 不 可 大 也
bí dà kě xiăo xiăo bù kě dà yě
nose big can small small NEG can big EQ
‘If a nose is big, one can make it small, if it is small, one cannot make it big.’ (Bisang 2013:280)

In the following example, chén ‘minister’ in the transitive construction has the interpretation ‘be

15 Intransitive verbs have two subtypes: [+control] and [-control]. A [-control] verb will have CAUSE and BECOME operators, while a [+control] verb will have only a CAUSE operator. See Bisang 2010 for a detail.
like a minister’:

(4.7) 然則臣王乎
rán zé chén wáng hū
be.so then v:behave.like.a.minister king Q
‘Since this is so, will you serve the king as a minister?’ (Bisang 2013:282)

Therefore, the verb position in both intransitive and transitive structures allows not only prototypical action-denoting lexemes, but also object-denoting and property-denoting lexemes. However, the fixed argument positions signal the grammatical function of the lexemes in a sentence without category specification, and the appropriate interpretation is possible from the meaning attributed by the construction and the lexeme itself. For instance, the meanings of object-denoting lexemes in the verbal slot are: (1) human beings, (2) instruments/man-made objects, (3) sense organs, (4) places and buildings, (5) pronouns of 1st and 2nd person, and (5) numbers and measures. If a ‘human being’ appears in an intransitive construction as in (4.5), the interpretation will be that the NP behaves like a (true) N, and if it appears in a transitive construction as in (4.7), it will be interpreted to mean that the NP\(_A\) causes the NP\(_U\) to be/behave like a (true) person-denoting lexeme.

Examples (4.5-4.7) show that Late Archaic Chinese does not have any morphological grammatical markings on the lexical items, but Bisang explains that Old Chinese\(^{16}\) did have a prefix *s- and a suffix *-s that attached to bases in both N-position and V-position. Therefore, at first glance, the affixes can attach to any base, which means that the morphology is not category sensitive. However, once an affix is attached, the derived word is always associated with a particular word class. If the same affix attaches to another base, it can derive a word that belongs

---

\(^{16}\) Late Archaic Chinese (fifth to third century BC) represents the last stage of Old Chinese (eleventh to third century BC) (Bisang 2010, 2013).
to a different word class. Thus, in Old Chinese, morphologically marked lexical items were
assigned to specific word classes, even if the affixes themselves were not category-specific.
However, Late Archaic Chinese lost the Old Chinese affixes, and what remains is the argument
structures, that is, the syntactic slots. Therefore, Bisang argues that Late Archaic Chinese has
syntactic categories, but not lexical categories, and the lexical items are “coerced” into acquiring
a syntactic category N or V.

Foley (1998) and Bisang (2010, 2013) claim that the Philippine languages and Late
Archaic Chinese do not possess lexical categories, but do have syntactic categories N and V. In
the next section, I will introduce an analysis that rejects not only the lexical categories but also
the noun and verb distinction.

4.5. Broschart (Tongan)

Broschart’s (1997) account of Tongan word classification is truly remarkable in the sense
that it broke the traditional, well-accepted, and intuitively appealing relationships between verbs
and predicability and nouns and referentiality. These traditional relationships are encapsulated in
Hengeveld’s (1992:58) famous proposal that “a verb is a lexical unit which, without further
measures being taken, has a predicative use only” and “a noun is a lexical element which,
without further measures being taken, can be used as the head of a term.” This concept has deep
origins in Greek ονόμα ‘name’ and ρήμε ‘utterance’, which correspond to the modern terms
“noun/verb” and possibly “subject/predicate.” In spite of theoretical and descriptive differences
in how to account for so-called “unfamiliar” part-of-speech systems, the relationship between the
noun/verb distinction and referentiality (DET phrase)/predicability (TAM phrase) was not
contested. Although some languages may not differentiate nouns and verbs in the same manner
as Latin or English do, there has never been a language discovered in which nouns are used primarily for predicates and verbs for reference. If there were, it would go against Croft’s (1991, 2001, 2004) prototype typology of parts of speech. The dichotomy of noun for reference and verb for predicate was an unshakable and well-established premise, until Broschart argued that it is not the case for Tongan.

Tongan is known to have flexible word class, as shown below:

(4.8)  
\[
\begin{array}{ll}
\text{te} & \text{u} \\
\text{puaka} & \\
\text{FUT} & \text{ISG} \\
\text{pig/pork} & \\
\end{array}
\]  
‘I will provide pork.’ (Broschart 1997:135)

(4.9)  
\[
\begin{array}{llllllll}
\text{ko} & \text{e} & \text{‘alu} & \text{‘a} & \text{Sione} & \text{ki} & \text{kolo} \\
\text{PRST} & \text{DET} & \text{go} & \text{AL.Gen} & \text{Sione} & \text{ALL} & \text{town}^{17} \\
\end{array}
\]  
‘Sione is going to town.’ (visibly for the hearer) (Lit. ‘It is going of Sione to town.’) (Broschart 1997:152)

In Tongan, a so-called noun *puaka* ‘pig/pork’ can follow the TAM and a so-called verb *‘alu* ‘go’ can follow the DET (ART in Broschart’s term). This flexibility is not without constraints.

(4.10)  
\[
\begin{array}{llllllll}
\text{lea} & \text{faka-Tonga} & \text{mai!} \\
\text{SPEAK} & \text{‘ADV’-Tonga} & \text{to me} \\
\text{‘Speak to me in Tongan!’} & \text{(Broschart 1997:132)} \\
\end{array}
\]

The above example would not be acceptable without the “adverbializer” *faka*.

And although a so-called noun can follow TAM, TAM cannot take the ART-marked phrase (ART-syntagm in Broschart’s term):

(4.11)  
\[
\begin{array}{llllllll}
\text{na’e} & \text{fefine} & \text{kotoa} & \text{e} & \text{kau} & \text{lelé} \\
\text{TAM} & \text{woman} & \text{all} & \text{DET} & \text{NUMBER} & \text{run} \\
\end{array}
\]  
‘The ones running were all female.’ (Broschart 1997:134)

(4.12)  
\[
\begin{array}{llllllll}
*\text{na’e} & \text{e} & \text{tangatá} \\
\text{TAM} & \text{DET} & \text{man} \\
\end{array}
\]  
‘It was a man.’ (Broschart 1997:133)

---

17 Broschart’s abbreviations: ADVL adverbializer; AL alienable; ALL allative case.
Therefore, not all lexical items can appear freely in the TAM slot and the DET (ART) slot. Because, prototypically, TAM is used with a verb for predication and DET is used with a noun for reference, the most likely scenario is that Tongan has rampant zero-derivation, which derives a verb \textit{puaka} ‘(to) provide pig/pork’ in the TAM phrase in (4.8) from a noun \textit{puaka} ‘pig/pork’. But Broschart claims that this is not a case of zero-derivation. He defines zero-derivation thus: “certain environments automatically change the meaning of the item they combine with, even if the base form does not change” (Broschart 1997:135), and claims that the meaning of \textit{puaka} does not change whether it is in a TAM phrase or a DET phrase, although the English translation shows otherwise (\textit{puaka} [n] ‘pig/pork’: \textit{puaka} [v] ‘provide pig/pork’). Tongan allows even the interrogative pronouns such as \textit{hā} ‘what’ to be used in both TAM and DET phrases, and it is extremely difficult to interpret \textit{hā} ‘what’ in a TAM phrase as having a “verbal” meaning or being used “verbally.” Therefore, Broschart raises this issue: The TAM-syntagm and the ART-syntagm have different functions, which is expected, but because Tongan ignores this distinction and allows many lexical items to appear in both syntagms, there must be something that TAM- and ART-syntagm have in common.

Broschart claims that there are two types of languages: nominal/verbal languages, such as Latin, and type/token languages, such as Tongan. The differences are presented in the following schemas:
For type/token languages, the most dominant factor in categorization is not [+/- pred], it is [+/- ref], which makes the distinction between the lexical items TYPE and the syntactic items TOKEN. Both TAM and ART phrases in Tongan are referentially marked in discourse, unlike the Latin or English type of languages, in which [+ref] goes only with N(P). A syntactic “token”
is referentially marked, and Broschart argues that the feature [+ref] is what the TAM phrase and the ART phrase have in common.

(4.13) ‘oku laione pē ‘a e laioné
   TAM lion just ABS DET lion.DEF
   ‘Lions will be lions.’ (lit. ‘The lion is just (like a) lion.’)

(4.14) ko e laione ia na’e ‘asi mai ‘i he viteó
   PRST DET lion 3SG PST appear to.me LOC DET video.DEF
   ‘It is/was a lion I saw on the video.’ (Broschart 1997:159)

He states that, in (4.14), only laione in the ART phrase is referential, as it is referred to anaphorically, while laione in the TAM phrase is not. However, Broschart claims that a TAM phrase as a whole can have [+ref] characteristics, because the whole phrase can be relativized in discourse, such as na’e ‘uha ‘it rained’, which one can refer to and say, ‘that was bad’. Unlike the nominal/verbal languages in which [+/- pred] has primary dominance in category distinction and makes the distinction between verb and noun, in Tongan, [+/- pred] has secondary dominance, and belongs to the phrasal level. A TAM phrase with [+pred] is a syntactic category IP and an ART-phrase with [-pred] is DP. Thus, Tongan grammatical categorization can be done without referring to the traditional terminology of noun and verb.

Broschart is not claiming that content words in Tongan are category-less. Rather, he suggests that Tongan words are category-less in terms of the noun/verb distinction. Tongan words have two primary categories: type and token.

Broschart rejects the classic part-of-speech system for Tongan, and Peterson does the same in his analysis of Kharia, presented in section 4.6.

4.6. Peterson (Kharia)

Kharia is a Munda language (Austroasiatic), spoken in eastern-central India, and is
known as one of the languages that do not show a noun/verb/adjective distinction. Kharia is similar to Mundari, which is also a Munda language that is assumed to be lacking such lexical classes. Evans and Osada (2005), applying their three criteria of compositionality, bidirectional flexibility, and exhaustiveness through the lexicon, argue that Mundari does have the noun/verb/adjective distinction, as discussed in Chapter 2. Peterson’s (2013) approach is unique in the sense that he argues that the analysis of such languages should be not only descriptively adequate but also the most economical—he asks: What does it buy to impose the traditional part-of-speech system on these languages? Peterson concludes that, for Kharia, an analysis without these lexical categories is better.

Kharia is a predicate-final language with grammatical categories such as person, number, tense, and case realized by enclitics and postpositions. Peterson (2013) divides the Kharia lexicon into contentive and functional morphemes. Contentive morphemes are used to form content heads; they can be used predicatively, referentially, and attributively; and they serve as a semantic base, while the functional morphemes are used in functional heads and for marking grammatical functions. In the following example, the contentive morphemes rochoʔb ‘side’ and col ‘go’ form content heads, while the grammatical categories =te and =ki=ɲ are cliticized as oblique marker and tense and person marker, respectively.

(4.15) ho rochoʔb=te col=ki=ɲ
     that side=OBL go=MV\(^{18}\).PST=1SG
     ‘I went to that side.’ (Peterson 2013:134)

Kharia is regarded as a flexible language because ho rochoʔb ‘that side’ can be used as contentive morpheme in a TAM/Person-syntagma:

\(^{18}\) MV stands for middle voice.
(4.16) ho\textsuperscript{19} rocho\textsuperscript{2} b=ki=n
that side=MV.PST=1SG
‘I went to that side.’ (lit: ‘I that-side-d.’) (Peterson 2013:134)

In (4.15), rocho\textsuperscript{3} b ‘side’ in rocho\textsuperscript{3} b=te ‘to (that) side’ is the content head, and the clitic =te is the
functional head, while col ‘go’ in col=ki=n is the content head, and =ki=n is the functional head.
The cliticized functional morphemes are attached to the end of the syntactic unit. rocho\textsuperscript{3} b=te is a
Case-syntagma, also referred as the “complement” (arguments and adjuncts), and col=ki=n is
the TAM/Person-syntagma, or “predicate.” A Kharia clause has to be either a Case- or
TAM/Person-syntagma, but it cannot be both. In other words, a “case\textsuperscript{20}”-marked item cannot
follow TAM/person marking. Therefore, a string of morphemes such as *city-OBL-MV-1sg ‘I
went to the city’ is not grammatical. Both Case- and TAM/Person-syntagmas consist of two “co-
heads”: a Content head and a Functional head.

Peterson (2013) shows the structures of Case-syntagma and TAM/Person-syntagma as
follows:

(4.17) Case-syntagma

(LEX=GEN)(DEM)(QUANTP)(LEX=GEN)(LEX*) (=POSS)(=NUM)=CASE

(4.18) TAM/Person-syntagma

CONTHEAD (V2*\textsuperscript{21}) (=PERF)=TAM/VOICE=PERS[NUM]/HON\textsuperscript{22}

Note that in (4.18), the CONTHEAD (CONTENT HEAD) is everything to the left of CASE.
LEX* means that any number of lexical items can appear in this slot, and LEX=GEN is the

\textsuperscript{19} In a later analysis, ho ‘that’ is labeled as DEM, which makes up the CONTENTIVE HEAD with rocho\textsuperscript{3} b ‘side’.  
(Peterson 2013:154)
\textsuperscript{20} Peterson (2013) includes zero-marked Direct case (for subjects and indefinite direct objects for 3SG), Oblique
case, and postpositions for “Case” in Case-syntagma. Genitive case is not included.
\textsuperscript{21} V2s refer to Aktionart, marking telicity and conativity, the passive, benefactive, etc.
\textsuperscript{22} HON stands for honorific.
genitive attribute. Case-syntagma and TAM/Person-syntagma are mutually exclusive, as they can never appear together, and the functional heads clearly mark the types of syntagma. The phrase-structure rules can be drawn for both syntagmas to account for Kharia sentences without referring to a noun/verb/adjective distinction.

Peterson’s (2013) claim is that the traditional parts of speech can be made to fit an account for Kharia word categorization by positing zero-derivation, but what he sees is that Kharia has two syntactic types, one marked by “cases” and another by TAM, which he calls “clear POSITIVE evidence” (Peterson 2013:145). Therefore, the primary grammatical category in Kharia belongs to the syntactic level, not to the lexical level. He adds that Kharia does not appear to be completely flexible, but if a certain semantic base is restricted from appearing in a particular syntagma, it is most likely because that particular base is not semantically compatible with the function specified in the syntagma. As such cases can be stored in the lexicon as ones that are not flexible, they are not a hindrance to his analysis of Kharia without the traditional part-of-speech system. He concludes:

Thus, by concentrating our discussion on whether or not a particular language can indeed be shown to have at least one noun and/or verb, or that these categories CAN be made to fit a particular language (with some ingenuity), we are probably overlooking quite a bit of other, potentially much more interesting, information on languages of this type. (Peterson 2013:168)

This insight is shared by Gil, whose study on Riau Indonesian is discussed next.

23 See Peterson 2013 (151) for more detail.
4.7. Gil (Riau Indonesian)

Riau Indonesian is a variety of Malay/Indonesian spoken in informal situations. According to Gil (2000, 2013), Riau Indonesian has only one syntactic category. *Makan ayam* ‘eat chicken’ and *ayam makan* ‘chicken eat’ can both mean ‘the chicken is eating’. Both *makan* ‘eat’ and *ayam* ‘chicken’ have the same syntactic behavior and same distributional properties and thus belong to the one and only category S. There is no head in this sentence, and it simply signifies ‘something about eating and chicken’. When the head is found in the sentence, an interpretation is given accordingly. Therefore, when ‘eat’ is the head in the sentence ‘eat chicken’, the sentence is talking about ‘eating associated with chicken’. A possible translation might be ‘Someone is eating the chicken’, or, equally possible, ‘The chicken is eating’ or ‘He ate it with/because of the chicken’. On the other hand, if ‘chicken’ is the head in ‘eat chicken’, then the meaning will be ‘chicken associated with eating’. In this case, the translation might be ‘the chicken which ate… or ‘the chicken which he ate…’, and so forth. Provided that Gil’s description is correct, it is apparent that traditional word classification and a grammar based on such word classification would not be useful for Riau Indonesian. With all these possibilities for different interpretations, it is apparent that thematic roles do not play a big part in Riau Indonesian grammar. There is no case-marking and no morphological agreement, and the word order is flexible in Riau Indonesian. Gil tested Riau Indonesian data using the three criteria put forth by Evans and Osada (2005), bidirectionality, compositionality, and exhaustiveness, and also applied ten criteria to test the noun and verb distinction. The results demonstrated that Riau Indonesian does not differentiate between nouns and verbs. Not only that, but Gil concludes

---

See Gil 2013 for more detail.
that Riau Indonesian does not possess lexical categories such as noun, verb, adjective, or any distinction between lexical and phrasal categories, or any distinction between lexical and functional categories. There is only a single primitive syntactic category $S^0$ in Riau Indonesian.

Gil (2013) questions why it attracts so much attention when a language lacks the traditional part-of-speech system, when no linguist questions why English does not have a grammatical category of dual number or morphosyntactically distinguish alienable and inalienable possessions. The traditional part-of-speech system has become the norm and the default hypothesis, which, according to Gil, is counter to Occam’s razor. He argues that the flexible system should be the default, because it is simpler, and “syntactic categories should be assumed to be absent until positive evidence is found” (2013:90). He proposes a type of language with the simplest word class system, which he calls Isolating-Monocategorial-Associational (IMA) language, and he provides the following definitions (2013:125):

(4.19)  

a. **Morphologically Isolating**

   No word-internal morphological structure

b. **Syntactically Monocategorial**

   No distinct syntactic categories

c. **Semantically Associational**

   No distinct construction-specific rules of semantic interpretation (instead, compositional semantics relies exclusively on the Association Operator)

Gil argues that, although no language fully meets the three criteria of an IMA language, languages show different degrees of simplicity/complexity in the three criteria, and that Riau

---

25 Gil (2013:125) claims that monocategoriality can be witnessed in apes’ communication and also in the language of young infants.
Indonesian is definitely the one language known at this point that shows the greatest degree of simplicity in all three.\textsuperscript{26} Thus, Riau Indonesian ranks lowest (that is, simplest) in the typology of IMA language. Gil further claims that Riau Indonesian is not an extreme and isolated case with extremely simple grammar, as other languages such as colloquial Malay/Indonesian dialects show similar grammars.

In the next section, I will discuss Biggs’s account of Māori word classification. Although Biggs was working several decades earlier and his goal was the description of Māori grammar rather than its analysis, his work and the analyses of the languages presented in sections 4.3 to 4.7 share a common understanding of languages with flexible word classes.

4.8. Biggs (Māori)

In chapter 3, I introduced the word classification system Biggs proposed for Māori. He claimed that phrases are the primary units for Māori grammar, and he classified lexical items (that appear in the nucleus) into categories by their compatibility with the phrase markers, TAM and DET. This classification method eliminated the verb and adjective classes, narrowed the noun class, and established a hugely conflated class, “universal.” For Biggs, the multifunctional nature of most of the Māori words needed to be addressed; he saw it as meaningless to label a word like pai ‘good’ a noun, verb, and adjective; classification in this situation becomes futile. Although Biggs mentioned that the inadequacy of the conventional part-of-speech system when it is applied to Māori is the reason that he established a new classification system to avoid

\begin{itemize}
\item This claim suggests that the existence of languages such as Riau Indonesian, which satisfies almost all three criteria for IMA language, refutes Frujyngier and Shay’s proposal that category underspecification at one level of grammar is counterbalanced at another level.
\end{itemize}
category overlapping, he does not claim that Māori lacks categories.

As explained earlier, Broschart, on the other hand, analyzes a similar language, Tongan, as a language to which the conventional part-of-speech system does not apply. Massam (2005) also agrees with Broschart that some languages possess a different word classification system from that of English. She uses Biggs’s category system, of nouns and universals\(^\text{27}\) to analyze Niuean (Polynesian). However, Massam argues that Niuean has categorical distinction at the lexical level. She posits functional heads (light v and light n) in Niuean, and, from the distributional evidence—unlike Tongan, not all Niuean words that are intuitively nouns can appear freely as predicates—Massam claims that there is category selection based on the lexical categories. Niuean has two classes of words that can appear with light v. In contrast to nominals that cannot appear freely as predicates, words that are intuitively verbs can appear rather freely as nominals, which means that light n can take both categories. In addition, in Niuean pseudo noun incorporation, words that are usually classified as verbs can appear in the same position in which nouns appear. Massam suggests, based on such evidence, that Niuean verbs are not verbs on the morphosyntactic level. This is similar to Biggs’s analysis of Māori in the way that he considers a nominal and verbal distinction to exist on the phrasal level, the former marked by DET and the latter marked by TAM, while there is no class “verb” at the lexical level.

Therefore, although Biggs did not claim that Māori does not differentiate categories at the lexical level, the facts that he pointed out may lead to a possible account of Māori as a lexically precategorial language.

\(^{27}\) Massam (2005) claims there is no word class in Niuean equivalent to Biggs’s “stative” in Māori. Therefore, her classification includes only “universals.”
4.9. Conclusion

In this chapter, I have introduced theories that do not consider lexical categories to be an inherent part of lexical items. Even within the framework of generative grammar, precategoriality was adopted in distributed morphology and similar models, which claim that lexical items are not given categories such as N or V a priori, but, instead, there is a mechanism that assigns the categories. In the distributed morphology framework, the same syntactic rules that apply to the word derivation system apply to the roots, which are not specified for categories. Lieber shows that the semantic information, including the argument structure, of affixes and stems can account for affixal operations without referring to categories. Baker claims that it is the syntactic structure that assigns category. These linguists do not deny the traditional part-of-speech distinction; what they reject is the idea of inherently given lexical categories, which implies that the category relevant to grammatical theory is the syntactic category.

Lexical underspecification theories have been applied especially when the languages under investigation are considered to have “flexible” category. Foley (1998) proposes that Tagalog does not posses lexically specified word classes, judging from the fact that, first of all, Tagalog so-called “verbs” are never used without the overt verbal affix that corresponds with the NP that serves as the pivot of the sentence. NPs that are considered oblique, such as instrument, can serve as the pivot in Tagalog. This indicates that the “verbs” in Tagalog do not have argument structures in the conventional understanding, such as transitive verbs with Agent and Patient arguments, which disqualifies them as “verbs.” As the Tagalog “verbs” can also take the affixes that attach to “nouns,” Foley concludes that Tagalog does not differentiate noun and verb on the lexical level. Bisang (2010, 2013) makes the same claim about Late Archaic Chinese, and he explains it from the diachronic point of view. Old Chinese did show morphological markings
that signify word categories, but these were dropped in Late Archaic Chinese, leaving only the syntactic slots as the category determinants. He claims that the part-of-speech distinction cannot be an inherent property of individual lexical items if it can be lost so easily with the disappearance of its overt marking, while without the syntactic slot that assigns the grammatical functions to each word, the proper interpretation is impossible in Late Archaic Chinese. Therefore, the syntactic categories are of primary importance in Late Archaic Chinese.

Broschart (1997), Gil (2000, 2013), and Peterson (2013) not only claim that there are no categories on the lexical level, but also that categories do not have to fit the traditional part-of-speech system. They suggest that the overlapping of categories can be better addressed by not imposing the traditional part-of-speech system. Broschart showed that Tongan is a language that operates with an entirely different word classification system, in which elements are primarily categorized by the property [+/- reference], which sorts out lexical items [-ref] and syntactic items [+ ref]. He calls this a type/token language, claiming that the noun/verb distinction is not necessary to account for Tongan grammar. Peterson points out that it is not that a language cannot be explained with the traditional part-of-speech system, but that it is not necessary to make the traditional part-of-speech system fit when there are other alternatives. He showed that phrase-level categories—TAM/Person-syntagma and Case-syntagma—are of basic grammatical importance in Kharia, which, like Tongan, does not need the N/V distinction at the lexical level. Gil is against the practice of treating the traditional part-of-speech system as the default model. He presents Riau Indonesian as an example to show how “different” a language can be in terms of its word categorization, as Riau Indonesian, according to Gil, has only one type of syntactic category. Regardless of the differences in the theories discussed in this chapter, they share one common assumption: The categories (whether or not they fit the traditional part-of-speech
system) exist on the syntactic level (albeit only one in Riau Indonesian). The models introduced here show that the grammar of languages, and not just languages with flexible category but those like English as well, can be explained without lexical categories. In the next chapter, I will consider whether a precategorial analysis can be applied to Māori and if such an account is better suited for Māori.
Chapter 5

Māori as Phrase-based Language

5.1. Māori Phrases as Basic Grammatical Units

In the previous chapters, I have discussed the origin of the concept of parts of speech and its development in the study of language. I showed that the original part-of-speech system is a Greek-based, language-specific, primarily morphological and semantic word classification system, which is still used widely in various fields, including pedagogy and theoretical and descriptive linguistics. Parts of speech are described morphologically, notionally, and syntactically, according to the language under investigation. If a language lacks morphological markings to indicate parts of speech, then syntax or semantics provide the primary criteria. As more research has been done on non-Indo-European languages, the conventional part-of-speech system has been found to be inadequate. Because their behavior is puzzling if viewed in terms of the conventional noun/verb/adjective distinction, these “exotic” languages have given linguists a chance to question the foundations of the classic part-of-speech system and explore alternative explanations for word categorization.

In chapter 3, I showed that Māori words are very flexible in their selection of TAM and DET. A phrase marked with TAM is equivalent to IP in generative grammar, while a phrase marked with DET is equivalent to DP.28 In short, the tree structures for Māori phrases in Figure 5.1 need to be accounted for.

28 If the NP analysis is used, then [NP DET[N[N/V A N/V A]]] has to be accounted for.
Figure 5.1. Tree structure representations for Māori IP and DP

Linguists have found various ways to explain the mismatch between the phrasal category and the lexical category when a very “verby” word such as *haere* ‘go’ appears with a determiner or a very “nouny” word such as *rākau* ‘tree’ appears with TAM. Proposed explanations include multiple categories for individual words (e.g., *haere* [V], [N]); a new class (“universal”) of multifunctional words; and regularly occurring zero-derivation between categories (e.g., *haere* [V], *haere* [N]).

In spite of their differences of opinion on word categories, there is one thing that Māori grammarians all agree on: A phrase, not a word, is the basic unit of Māori grammar. Biggs (1961) introduced the term “contour word” for the phrase that serves as the “natural grammatical unit” (Biggs 1969:3) in Māori, both phonologically and syntactically. Although a phrase consists
of three parts—preposed periphery (TAM or DET), nucleus (bases/content words), and postposed periphery (optional)—this string is considered a minimal syntactic unit, as if it were a word (a “contour word”). While Māori reference grammars mention this fact first, they then carry on with business as usual, treating the individual word, not the phrase (or contour word), as the basic unit, including discussion of its part of speech. In the existing grammatical descriptions of Māori, the significance of the proposition that the phrase is the basic grammatical unit is not clear, nor is it clear whether and how that fact is relevant to various aspects of the grammar. Yet if a phrase, not a word, is the basic grammatical unit, the fact must have some ramifications for grammatical analysis.

For example, if a phrase is the basic grammatical unit, equivalent to a word, the three slots are a word-internal structure. Take the phrase *kua hoki mai* ‘(someone) has come back’, for example. As illustrated in Table 5.1, the TAM *kua* in the preposed periphery can be considered an inflectional morpheme, and the base can be considered a bound morpheme.

**Table 5.1. TAM-phrase as a word**

<table>
<thead>
<tr>
<th>Preposed periphery</th>
<th>Nucleus</th>
<th>Postposed periphery</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>kua</em> TAM</td>
<td><em>hoki</em></td>
<td><em>mai</em></td>
</tr>
<tr>
<td>VERB INFLECTION</td>
<td>BASE</td>
<td>DERIVATIONAL SUFFIX</td>
</tr>
</tbody>
</table>

The particle in the postposed periphery is optional. If it is present, it provides an additional description of the base. *Hoki atu* ‘return away’ means ‘go back’, as opposed to *hoki mai* ‘come back’. In that sense, it can be considered a derivational suffix, which is similar to the English prefix *re*- in *return*. The DET-phrase can also be analyzed as a word. There are languages like Swedish in which definiteness is expressed by affixation. Therefore, it is possible that DET is a
prefix for a noun declension.

**Table 5.2. DET-phrase as a word**

<table>
<thead>
<tr>
<th>Preposed periphery</th>
<th>Nucleus</th>
<th>Postposed periphery</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>te</em> DET</td>
<td><em>pukapuka</em> ‘book’</td>
<td><em>nei</em> ‘this’</td>
</tr>
</tbody>
</table>

NOUN DECLENSION STEM DERIVATIONAL SUFFIX

In this analysis, the categories of *hoki* ‘return’ and *pukapuka* ‘book’ are not lexical or syntactic, but are categories of bound morphemes.²⁹ According to this interpretation, Māori would possess both derivational and inflectional morphemes, and it could not be the lack of inflectional and derivational morphemes that causes category overlap.

However, an item like *hoki* or *pukapuka* is not usually taken to be less than a word. If they are not bound morphemes, but independent words (potentially free morphemes), and they are not basic grammatical units, then what roles do Māori words play in the grammar that are different from the roles of words in other languages? And what exactly does it mean to say that the phrases are the “natural grammatical units”? In this chapter, I will show that Māori phrases—DET-phr(ase) and TAM-phr(ase)—are the category carriers, and there are no lexical categories inherent to lexical items in Māori grammar. The lexical items that can appear in the nucleus, which have lexical meanings (the “bases”), are category-free, possess only the meaning (or rather, the concept), and have the potential to co-occur with either TAM or DET as long as the combined units are semantically compatible with each other. That is, if some bases select or reject certain types of TAM or DET markers, it is due to semantic constraints. In Māori, while TAM and DET form with their complements two distinct types of phrases, they share the same

²⁹ Before Biggs (1969) renamed these “particles and bases,” he called them “major and minor morphemes,” respectively (1961). Affixes such as -*Canga* and -*Cia* were “nuclear minor morphemes.”
distributional slot inside the phrase, and although [+/-pred] is what separates them, it is not as absolute and crucial a criterion in defining parts of speech as it is in English. I will point out that Croft’s (2001) prototype analysis does not prove useful for Māori, although it does superficially identify the prototype nouns, verbs, and adjectives, none of which is irrelevant to the syntax of Māori. I will first show how the noun/verb/adjective distinction can be applied to Māori to examine whether it is beneficial to the description of Māori grammar.

5.2. The Classic Part-of-Speech System and Māori Grammar

Despite the high degree of category overlapping in Māori and despite the acknowledgment of the phrase as the basic grammatical unit, most Māori reference grammars operate on the assumption that Māori words fit into noun, verb, and adjective categories. Biggs (1969) proposes the new word class “universal,” which includes all the items that can take passive forms. This lumped category includes words such as oma ‘run’, rākau ‘tree’, and pai ‘good’. This “grab bag” treatment of Māori words does not sit well in the minds of other Māori scholars, and therefore, while acknowledging that Māori exhibits the phenomenon of category overlapping, they adopt the classic part-of-speech system in their descriptions of Māori grammar.

Biggs (1961, 1969) posits the word classes of universal and noun, arguing that the verb as a word class does not exist. Table 5.3 shows Biggs’s classification system, with his examples (1969).
### Table 5.3. Nouns and verbs

| Nouns | Definition:  
|-------|----------------------------------------------------------|
|       | the bases that do not occur with TAM; do occur with the determiner te/ngā; do not occur with -Cia or the personal particle (e.g., a Hone ‘John’); and cannot immediately follow ki.  
|       | Examples:  
|        | rākau ‘tree’ *ka rākau *ki rākau  
|        | ngaru ‘wave’ *kei te nagru *ki ngaru  
|        | ika ‘fish’ *e ika ana *ki ika |
| Universals | Definition:  
|           | the bases that take passive -Cia  
| Examples: | inu ‘drink’ inu.mia ‘be drunk’  
|           | tangi ‘cry’ tangi.hia ‘be wept for’  
|           | kī ‘say’ kī.a ‘be said’ |

Of the three examples of nouns Biggs gave, the first two (rākau and ngaru) are in fact grammatical when used with TAM in certain contexts, according to some speakers.

Bauer (1993:254) comments on Biggs’ categories:

If this narrow definition is maintained, then the description fails to capture the fact that many but not all of Biggs’s class universals will behave like these nouns on particular occasions, while some universals never do. It also fails to capture the fact that there are meaning differences of a relevant kind associated with differences in syntactic usage for universals.

Bauer (1993) argues that there are items in the category of universals that share common characteristics of nouns, and that Biggs’s system, especially by establishing the class of universals, does not capture the fact that, although many words do co-occur with either TAM or
DET, they also show different behaviors. According to Bauer, “syntactic processes in Māori do refer to some of the traditional classes which Biggs discards” (Bauer 1993:254).

First, Bauer argues that the noun’s co-occurring with the determiner cannot be a reliable criterion, because there are exceptions. She gives the following example:

(5.1) …pau katoa i a ia te kai—kākahu atu, meremere atu; parawai atu; exhausted all cause PERS 3SG the food clothes away meremere away fine mats away

   aha atu; katoa ngā mea a te Māori…
   what away all the thing of the Māori

   ‘He finished all the food—clothes, merereres, fine mats, etc., all the things which a Māori possesses…’ (Bauer 1993:255)

As she herself admits, this is one of the “occasional exemptions,” and the nouns in Māori usually are used in a phrase with determiners.30 In example (5.1), it is possible that te that precedes kai might semantically or discoursally include the nouns that follow, as the word atu, which is glossed as ‘away’ is used after each noun. In a more common usage, atu is used with me ‘with’ to imply that the noun is included:

(5.2) I whariru ia ki ngā tāngata katoa, me Hone atu
   TAM shake 3SG to the people all with John away

   ‘He shook hands with everyone, including John.’ (Bauer 1993:305)

30 Another exception is object incorporation, which, as example (i) shows, allows a noun to follow immediately after the verb, without the DO marker i.

(i) E hī ika ana ia
   TAM pull fish TAM 3SG
   ‘He is fishing.’

In object incorporation, the incorporated noun is no longer a direct object, as one can see by the loss of the DO marker and the DET in (i), both of which are present in the canonical transitive construction in (ii).

(ii) E hī ana ia i te ika
    TAM pull TAM 3sg DO DET fish
    ‘He is fishing.’
It appears that the items without determiners in (5.1) are additional items added to a partial list of the things that were exhausted or finished, that is, ‘all the things a Māori possesses’. Such examples are not crucial enough to deny the fact that nouns take determiners.

Bauer (1993) goes on to criticize the use of determiners, especially *te*, as the criterion for identifying nouns. She claims that the determiner *te* can appear in “constructions which are principally verbal” (Bauer 1993:255), as in *kei te* and *i te* for TAM and *ki te* for the infinitival purpose phrase, “where the head is followed by verbal arguments rather than by typically nominal adjuncts” (Bauer 1993:255). The infinitival purpose phrase uses *ki te*, as shown in example (5.3).

(5.3) Ka haere a Mere ki te moana *ki te* hī ika
TAM go PERS Mere to DET sea *to DET* pull fish
‘Mary went to the sea to catch fish.’

Bauer suggests that the indefinite determiners *tētahi/ētahi* can identify nouns more accurately. She claims that some of the bases that Biggs calls “universals” can appear in the slot in the following construction (5.4), while others cannot.

(5.4) Kāore au i te mōhio he aha ______ māku
NEG 1sg TAM know DET what DET ______ for me
‘I don’t know what to ______.’ (Bauer 1997:72)

According to Bauer, bases such as *horoi* ‘clean’, *hanga* ‘build’, and *āwhina* ‘help’ cannot appear in the slot, while *waiata* ‘sing’, *kai* ‘eat’, *whakaaro* ‘think’, and *pātai* ‘ask’ are acceptable. She further claims that the accepted bases can co-occur with *tētahi/ētahi*, while the rejected bases

________________________

31 Bauer does not explain why *horoi/hanga/āwhina* are not acceptable in the given construction. According to Dr. Wiri, with *horoi/hanga/āwhina*, the sentence “does not make sense.” The correct way to say this is Kāore au i te mōhio he aha tētahi mea hei horoi/hanga/āwhina māku (NEG 1sg P DET know DET thing wash/build/help for me) ‘I don’t know what it is for washing/building/helping’. He explained that, in this context, *waiata/kai/whakaaro/pātai* are “nouns” meaning ‘song/food/thought/question’, while *horoi/hanga/āwhina* are actions that do not fit in this context. My informant’s reply proves Bauer’s point: *waiata/kai/whakaaro/pātai* are nouns. However, my informant adds, “in this context,” because *horoi/hanga/āwhina* do have nominal usages such as *te hanga* ‘the business, the work’, etc.
cannot. She concludes that “only those stems which are both nouns and verbs can occur in this construction. Without these classes to refer to, it is not possible to specify which forms are possible in this construction” (1993:255).

It is not very clear what Bauer is trying to prove by this argument. She is against the category “universal” proposed by Biggs, because it simply holds everything that takes passive -Cia, whereas Biggs’s category “nouns” has too narrow a criterion, which selects only some bases as nouns. She finds that the bases that fit the universal category show very differing behaviors, while they also share similarities with bases that do not fit the universal category. She considers the indefinite determiners tētahi/ētahi to be more suitable than definite te to establish nounhood, because the indefinite determiners identify sing, eat, think, and ask as nouns, but not clean, build, and wash. However, her screening sentence in (5.4) identifies bases that are both nouns and verbs, not just nouns. She may have succeeded in disqualifying several bases from the universal category, but the fact remains that there are bases that have multiple categories, exactly as her test picks out. Bauer criticizes Biggs by stating that “Biggs is able to claim that no classes overlap because he has created the class ‘universal’ to contain essentially those items which would otherwise belong to overlapping classes” (1981:28). This criticism is unwarranted, as the class “universal” is the result of Biggs’s attempt to classify the Māori words without overlap, by, in a sense, accepting overlap as a defining characteristic of Māori. At the same time, this criticism illustrates the circular nature of the debate on word classes. A researcher searches for the solution to the problem of category overlap by moving beyond preconceived ideas and looking at the language with an unbiased view. If she finds, after this investigation, the necessity of positing a new type of category that ends up including all the words with multiple categories, this solution is criticized for not solving the overlap problem. Simply put, if a
noun/verb/adjective distinction is not posited, there is no category overlap.

However, Māori linguists such as Bauer (1993, 1997), Harlow (2001, 2007), Hohepa (1967), and Waite (1994) retain the traditional parts of speech in their analyses. If the part-of-speech system is applied, and the bases are classified as nouns and verbs (and adjectives) regardless of overlap, we arrive at the tree diagrams in figure 5.1, in which NP and AP can be the complement of I, and DP and AP can be the complement of Det, or V or A can be the head of NP. In the generative framework, it is assumed that D is the head of DP, which takes NP in its complement position \([D \rightarrow D \text{ NP}]\), and I is the head of IP, which takes VP for its complement \([I \rightarrow I \text{ VP}]\). The discrepancy we can see in the Māori trees in figure 5.1 between the phrasal category and the head of the lexical phrase in the complement (\([D \rightarrow D \text{ AP/VP}] [I \rightarrow I \text{ AP/DP}]\)) needs to be resolved either by allowing multiple categories for a single lexical item or by taking zero-derivation to be the default word formation process in Māori. Another solution is to make a language-specific stipulation in the theory to accommodate the Māori facts. Waite (1994) chooses this solution. I will examine his analysis of Māori word categories as an example of the noun/verb/adjective analysis.

In the generative framework he uses, Waite (1994) suggests that Māori functional categories (Infl and D) can take NP, VP, and AP for their complements. The phrase structure tree in figure 5.2 represents the structures in his analysis.
Figure 5.2. Māori phrase structure (Waite 1994)

According to Waite, there are two types of derivation: lexical and syntactic. Koroua ‘old man’ (n) can derive a verb to mean ‘to become an old man’ through zero-derivation, which is a lexical rule. Such a process of category change involves semantic changes, as between waiata ‘to sing’ (v) and waiata ‘song’ (n) in example (5.5).

(5.5) I whakaputina e te kaitito āna [N waiata] katoa
   TAM appear.PASS by DET composer POSS 3SG song all
   ‘The composer published all his songs.’ (Waite 1994:60)

Waite claims that a syntactic rule allows Māori nouns, verbs, and adjectives to appear in the determiner phrase without category change or semantic shift:

(5.6) Ka pakaru te wini i te waiata a te wahine
   TAM broken DET window by DET sing of DET woman
   ‘The woman’s singing broke the window.’

To clarify, Waite suggests that te waiata in (5.5) is a lexically derived noun ‘song’, while te waiata in (5.6) is a verb ‘sing’ that can appear in DP. He notes that “the direction of derivation suggested here for zero-affixation is intuitive only, and has no bearing on the arguments put forward” (1994:60 fn). He claims that he instinctively knows that koroua (v) ‘to become an old man’ is derived from koroua (n) ‘old man’, and not vice versa, but that tangi (n) ‘sound’ is derived from tangi (v) ‘to make sound’. That one can make instinctive decisions on the direction
of derivation also implies that one knows instinctively that the default category of *koroua* is noun. Therefore, Waite suggests that speakers know the word categories by instinct.

Waite defends making just one stipulation to justify the representation in the trees above, that is, allowing D to take multiple categories as complements, because it can “account for the facts of Māori more elegantly than previous studies” (1994:55). That is, one stipulation in the higher node allows the lexical items in the complements to maintain their categories—N, V, and A remain the heads of NP, VP, and AP respectively—while it also proves the advantage of DP analysis over NP analysis, showing the parallel between DP and IP. The alternative solution would be to postulate zero-derivation to convert all the categories so they can occur in the appropriate nodes, which Waite prefers to avoid because he considers such a rule to be redundant. By positing zero-derivation as a lexical derivation rule and by stipulating the syntactic rule that allows Māori functional categories to take NP/VP/AP as their complements, Waite maintains that Māori grammar can be accounted for simply. Waite (1994:61–62) offers the sentences in (5.7) to demonstrate that Māori DP can take NP or VP or AP as its complement.

(5.7)  

a. *te whare*

[DP [D’ te [NP whare]]]  
‘the house’

b. *te tere (ki) te whakahoki mai*

[DP [D’ te [AP tere (Dat)/te Caus.go back hither]]]  
‘the speed of reply’

c. *te kimi (i) ngā kupu hou*

[DP [D’ te [VP search.for/(Acc)/te word/new]]]  
‘the search for new words’

Unlike Bauer, who rejects Biggs’s method of establishing a new category (“universal”) for the bases that correspond to multiple categories, Waite finds a way, while still keeping the traditional parts of speech, to accommodate overlap in the theoretical framework of his choice.
However, Waite’s approach seems costly, because the stipulation he makes is rather serious in terms of the principles of X-bar theory. On one hand, he strictly observes the rules *[\text{NP} A]* and *[\text{NP} V]*, assuming that the head and phrase must be of the same category. On the other hand, he accepts *[DP [D’ DET, VP]]* or *[DP [D’ DET, AP]]*. Waite claims that the tree structure on the right in figure 5.3 (DP analysis) is preferred because the phrase structure is maintained with consistent category notation (DP-D’-D), whereas the tree on the left in the figure shows inconsistency in the categories within the phrase structure, since V and A cannot be the head of X.

![Figure 5.3. NP analysis (left) and DP analysis (right) for Māori DET-phrase](image)

In generative theory, because N, V, and A are abstract categories, each language has its own system to make correspondences between the actual lexical items and the abstract categories. Therefore, Waite suggests that the parameterization in Māori is *X = N, V, A*, where *X* can be the head of XP in the complement position in DP, because V and A cannot be the head of NP. The question is whether it is acceptable to assume DET + VP and DET + AP, simply because DET and V/A are in a mother-daughter relationship, but not DET + V and DET + A, a sister relationship, when the outcome is the same string (e.g., *te kaha* ‘DET strong’, *te oma* ‘DET
run’). It is not entirely clear how to determine that one tree is more acceptable than the other, because both accounts (NP analysis and DP analysis) violate the phrase structure rules in the theory. Waite retains the traditional parts of speech, N, V, and A, claiming that this choice is theoretically motivated and thus there is no need to abandon the categories in Māori because of the phenomenon of category overlap. However, the theoretical motivation is that, within the framework of his choice, the phrase structure rules are stated in terms of these features as syntactic primitives, by the notations N, V, A. As mentioned in chapter 2, generative grammar, at least the mainstream theories, was developed without digging deep into the definitions of the categories of lexical items. Waite’s account of Māori facts uses a theory with syntactic categories that are assumed without being well defined in the first place. Therefore, his account does not clarify anything about the overlapping nature of Māori word categories. In this sense, he simply presents Biggs’s universal category à la generative grammar.

5.3. Prototype Categories in Māori

As mentioned in the previous section, Waite (1994) claims that the direction of the zero-derivation presented in example (5.8) is intuitive.

(5.8) \textit{koroua} (n) ‘old man’ $\rightarrow$ \textit{koroua} (v) ‘to grow old (of a man)’

\textit{tangi} (v) ‘to produce a sound’ $\rightarrow$ \textit{tangi} (n) ‘(the resulting) sound’

He thus implies that \textit{koroua} ‘old man’ is underlyingly a noun, and \textit{tangi} a verb, which undergo zero-derivation for lexical derivation. At first glance, it seems rather difficult to accept the claim that a category decided by instinct can be used for formal linguistic analysis. While the choice of the word “instinct” is unfortunate, Waite’s intention can be explained by Croft’s (1990, 2001, 2004) prototype analysis, described in chapter 2, which is illustrated here again in table 5.4.
Table 5.4. Croft’s prototype categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Reference</th>
<th>Modification</th>
<th>Predication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objects</td>
<td>UNMARKED NOUNS</td>
<td>genitives, adjectivalizations, PPs on nouns</td>
<td>predicate nominals, copulas</td>
</tr>
<tr>
<td>Properties</td>
<td>deadjectival nouns</td>
<td>UNMARKED ADJECTIVES</td>
<td>predicate adjectives, copulas</td>
</tr>
<tr>
<td>Actions</td>
<td>action nominals, complements, infinitives, gerunds</td>
<td>participles, relative clauses</td>
<td>UNMARKED VERBS</td>
</tr>
</tbody>
</table>

As explained in chapter 2, prototype analysis categorizes the members of classes not by clear demarcation, but by identifying the core members that are the “best examples” of the class as the prototypes, placing the less likely members on a continuum. What Waite refers to as categories selected by instinct are prototype categories. Pai ‘good’ is prototypically an adjective as it denotes property and modifies nouns; waiata ‘sing’ is prototypically a verb, referring to actions and used for predication; koroua ‘old man’ is a noun, referring to an object (a person) and used for reference. Croft also predicts that all languages will have prototypical nouns, verbs, and adjectives that fulfill their primary propositional functions and their conceptual meaning, and which must have unmarked constructions. Māori is no exception. Therefore, it appears that Māori does possess items that fit the classic parts of speech, nouns, verbs, and adjectives. We will examine in section 5.6 whether or not the fact that some Māori bases fit the profiles of the

---

32 Japanese uses a great many Sino-Japanese words consisting of two Kanji, such as 創造 ‘create (wound + construct)’ and 豐富 ‘wealth’ (abundant + fortune). These words are all nouns regardless of their meanings, and require further measures to be used in predicates and as modifiers: 創造する (v) ‘create + do’ and 豐富な (adj) ‘wealthy (wealth + na)’. These examples are counter to Croft’s prototype definitions.

33 Harlow (2007: 110) claims that the noun par excellence in Māori is one with -Canga nominalization, most likely because -Canga derived nouns only select DET. This contradicts Croft’s proposal that the core members of a category (i.e., the prototypes) are unmarked.
prototype nouns, verbs, and adjectives means that these are the type of grammatical categories that best describe Māori grammar.

5.4. Consequences of Identifying Parts of Speech

Once a word is assigned to one of the traditional parts of speech, it leads to a certain direction in the analysis of the grammar of the language in question. For example, Waite’s (1994) understanding of waiata (n) ‘song’ as derived from waiata (v) ‘sing’ determines the direction of the derivation. Not only that, it becomes necessary to differentiate the types of derivation, as seen in (5.5) and (5.6) between te waiata ‘singing’ and te waiata ‘song’, the former being the category-changing lexical derivation, and the latter remaining a verb even though waiata ‘singing’ is marked by the determiner te (thus, it is a syntactic derivation like that of the English gerund). Waite also accommodates overlapping categories by making a stipulation in the theory to allow both functional categories, I and D, to take any categories, N, V, A, in the complement position. Next, the assignment of parts of speech also entails a certain understanding of suffixes. The -Canga suffix, for instance, is called a “nominalizing” suffix. The term strongly suggests that it is a category-changing derivational suffix, deriving a noun from a non-noun. However, as seen in chapter 3, although the derived items do indeed select only DET, -Canga can attach to a noun in order to derive an “abstract noun” and can also attach to an item whose function is to modify the preceding base with -Canga as agreement; the interrogatives can be “nominalized,” as well. Pēhea ‘how’ and aha ‘what’ can follow either TAM or DET (e.g., Kei te pēhea? (TAM how) ‘How are you?’ or I aha.tia tō waewae? (TAM what,pass your leg) ‘What happened to your leg?’, and He pēhea te nui o te tapatoru (DET how DET big of DET triangle) ‘How big is the triangle’ or he aha? (DET what) ‘why?’). When -Canga is attached, only DET
can be chosen (*te aha.tanga, but *ka aha.tanga). It is difficult to give an English gloss for the “nominalized” interrogative expressions. Describing -Canga with the term “nominalization” implies that pēhea and aha belong to a certain category, anything except a noun, so that nominalization can apply.

The so-called passive suffix -Cia attaches to bases that strictly select TAM. Therefore, similar to the way -Canga selects DET and thus identifies nouns, -Cia is known to identify verbs. However, although passivization is an operation on verbs, -Cia can attach to bases that are not verbs. As with -Canga and modifiers, a manner adverb after a base with -Cia also needs -Cia for agreement. It can attach to interrogatives, as in pēhea.tia and aha.tia, and to nouns, as in whare.tia and rākau.tia in kua whare.tia/rākau.tia te taone ‘the town is housed or tree-ed/the town is covered by the houses/trees’. Pai ‘good’, tika ‘correct’, and ora ‘well’ have passive forms as well. As it is inconceivable for nouns and adjectives to be passivized, either the passive analysis of -Cia is at fault or the category assignment is at fault.

In sum, although one might say that Māori has prototypical nouns/verbs/adjectives per Croft’s (1990, 2001, 2004) analysis, using these categories to account for Māori grammar creates inconsistencies, such as passive forms of nouns and adjectives, and noun forms of interrogatives. The necessary conclusion is that these are not the most appropriate classifications of Māori bases.

5.5. Facts of Māori

The following list recaps the facts of Māori.

1. A phrase (contour word) is the basic grammatical unit.

2. A Māori phrase is divided into three slots: a preposed periphery occupied by TAM or DET
particles, a nucleus occupied by bases (content words), and a postposed periphery occupied optionally by particles that have similar functions as adverbs.

3. There are some words that can appear independently, such as *katoa* ‘all’, unaccompanied by TAM or DET. But TAM and DET and other particles in the peripheries and the bases in the nuclei must appear as a unit, not independently (except in specific constructions such as imperatives).

4. TAM and DET are in complementary distribution.

5. Bases that can co-occur with TAM also co-occur with DET, while there are bases that can only take DET.

6. The two affixes -*Canga* and -*Cia* select only DET and TAM respectively.

Table 5.5 summarizes the phrase internal structure of Māori.

**Table 5.5. Māori tripartite phrase structure**

<table>
<thead>
<tr>
<th>PREPOSED PERIPHERY</th>
<th>NUCLEUS</th>
<th>POSTPOSED PERIPHERY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particles</td>
<td>Bases</td>
<td>Particles</td>
</tr>
<tr>
<td>TAM: <em>ka, kua, kei te,</em> etc.</td>
<td>semantic content items</td>
<td>rawa</td>
</tr>
<tr>
<td>DET: <em>te, ngā,</em> etc.</td>
<td><em>rākau</em> ‘tree’</td>
<td><em>atu</em></td>
</tr>
<tr>
<td></td>
<td><em>oma</em> ‘run’</td>
<td><em>mai</em></td>
</tr>
<tr>
<td></td>
<td><em>pai</em> ‘good’</td>
<td>(etc.)</td>
</tr>
<tr>
<td></td>
<td>(etc.)</td>
<td></td>
</tr>
</tbody>
</table>

As TAM and DET are in complementary distribution, the following combinations never occur:

(5.9) * TAM + DET-phr
* DET + TAM-phr
Also to be noted here is their compatibility with the prepositions. Māori does not possess many prepositions, thus each preposition has multiple functions. For instance, *i* is a past progressive TAM as in *i te*, a so-called direct object marker in transitive sentences and an indirect object marker in intransitive sentences, introduces location and time (i.e., translates as ‘at’), and is also used to mark the agent of a neuter verb. Setting aside the question of whether *i* with these different functions has different origins (as it is quite puzzling that one particle fulfills all these functions), I will call the particles that precede DET-phr “prepositions.” With this definition, the following observation is made:

(5.10)  
\[
\text{Prep + DET-phr} \\
\text{*Prep + TAM-phr}
\]

Harlow (2007:135–136) suggests that the nominal phrases (in Harlow’s terminology) are further divided into those with a preposition and those without a preposition. However, whether or not a preposition is required does not depend on the type of base in the phrase, but on the grammatical role the phrase plays in the sentence. Therefore, in this dissertation, I will not consider prepositions as having an independent status allowing them to form Prep-phr. The highly relevant aspect of prepositions for this study is that they can only introduce DET-phr in a sentence.

To discuss the parts of speech in Māori means to investigate the base categories in the nucleus. It is important to emphasize this point, because the individual word in the phrase is not supposed to be the primary unit of grammar. If there are still categories in the bases in the nuclei, such categories must be the kind of categories that contribute to the grammar of Māori, without which a successful account for Māori grammar is unattainable. In other words, we are not sorting the bases for any other reason than to accurately describe Māori grammar.
5.6. Base Category as a Phrase

Because there are only two options for the particles that the bases can follow, logically and ideally speaking, it must suffice to posit two classes for the base items in the nuclei: a class of items that co-occur only with TAM (Class A) and a class of items that co-occur only with DET (Class B). If the classic part-of-speech system and/or nouns, verbs, and adjectives nicely coincided with these two types of item, there would be no issue. However, as we have seen in chapter 3, that is not the case in Māori: The majority of the bases can take both TAM and DET, while there are some that take only DET. There are also affixes, such as -Cia for passive and -Canga for nominalization, which are more selective between TAM and DET. Table 5.6 shows actual Māori bases, rākau, pai, and kai, plugged into Croft’s table of the “overtly marked structural coding constructions for parts of speech” (2001:88).

Table 5.6. Parts of speech in Māori

<table>
<thead>
<tr>
<th>Reference</th>
<th>Modification</th>
<th>Predication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNMARKED NOUNS</td>
<td>genitives, adjectivalizations, PPs on nouns</td>
<td>predicate nominals, copulas</td>
</tr>
<tr>
<td>rākau ‘tree’</td>
<td>rākau ‘wooden’</td>
<td>rākau ‘is covered by trees’</td>
</tr>
<tr>
<td>Properties</td>
<td>deadjectival nouns</td>
<td>UNMARKED ADJECTIVES</td>
</tr>
<tr>
<td>pai/painga ‘goodness, being good’</td>
<td>pai ‘good’</td>
<td>pai ‘is good’</td>
</tr>
<tr>
<td>Actions</td>
<td>action nominals, complements, infinitives, gerunds</td>
<td>participles, relative clauses</td>
</tr>
<tr>
<td>kai ‘food/eating’</td>
<td>kai ‘of eating, of food’</td>
<td>kai ‘eat’</td>
</tr>
<tr>
<td>kainga ‘leftover’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

121
Note that, except for *painga* and *kainga*, which include the nominalizing suffix -*Canga*, the word form of the non-prototypes does not change. Modification can be done by simply placing a word after a modified noun, as in *whare kai* ‘dining hall’. Note that in English the pronominal modifier is marked; it is *dining*, not *dine*. Māori being a head-initial language, one understands the first base as the head and the second base as the modifier, even if the modifier does not have an extra morpheme to show its non-prototypical usage. The morphosyntactic measures that non-prototypes take to be used for non-prototypical usages need to be screened out, so that the discussion stays at the word level. Therefore, such measures as relative clauses, PPs on nouns, and complements need to be disregarded.  

For example, English *-ish* attaches to *book* (v) to make *bookish* (adj), which would nicely put *bookish* in the middle of the first row of the map as an example of adjectivalization. This would contain our discussion on the word level. We already know that Māori is not armed with such lists of category-changing morphemes to use in Croft’s table. Therefore, in Croft’s table, Māori prototype and non-prototype words are almost always identical in form. If this is all the map manifests for Māori, it means the map does not have much use for Māori and other languages like it.

However, if a phrase is the basic grammatical unit, possibly equivalent to a word, the preposed and postposed particles, and not only the bases, also must be mapped. Table 5.7 shows that *rākau*, *pai*, and *kai* can attach to the determiner *te* and to a TAM marker.

---

34 I will exclude the genitive from consideration, as the genitive is marked by the preposition *o/a*, which precedes the DET-phrase. Prepositions can be analyzed as part of the preposed periphery phonologically, but not syntactically.
Table 5.7. Māori phrases in a prototype table

<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
<th>Modification</th>
<th>Predication</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objects</strong></td>
<td><strong>UNMARKED NOUNS</strong></td>
<td><strong>rākau</strong></td>
<td><strong>kua rākau</strong> ‘has been covered by trees’</td>
</tr>
<tr>
<td></td>
<td><em>te rākau</em> ‘tree’</td>
<td><em>(pēne rākau)</em> ‘pencil’</td>
<td></td>
</tr>
<tr>
<td><strong>Properties</strong></td>
<td><strong>UNMARKED ADJECTIVES</strong></td>
<td><em>pai</em> ‘good’</td>
<td><strong>ka pai</strong> ‘is good’</td>
</tr>
<tr>
<td></td>
<td><em>te pai/te painga</em> ‘goodness, being good’</td>
<td><em>(te kōrero pai)</em> ‘the good talk’</td>
<td></td>
</tr>
<tr>
<td><strong>Actions</strong></td>
<td></td>
<td><strong>kai</strong> ‘of eating, of food’</td>
<td><strong>ka kai</strong> ‘eat’</td>
</tr>
<tr>
<td></td>
<td><em>te kai</em> ‘food/eating’</td>
<td><em>(te whare kai)</em> ‘the dining hall’</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.7 shows interesting patterns of the Māori bases. Of the three prototypes, only the prototype adjectives are unmarked, while prototype nouns and verbs are marked. At the same time, when object-denoting and action-denoting bases are used for modification, they are unmarked. These patterns have two possible implications: (1) Māori is an exception to Croft’s analysis, or (2) Croft’s analysis is wrong. The criticism might be made that the problem is an error in the input, that is, that the preposed particle should not be a part of the map, as the categories of the bases are the target of investigation. However, the simple fact is that the word forms of the bases alone do not tell their categories. Only when they are joined with preposed particles can the bases be placed in a semantic map, and this is in accordance with the premise that the basic grammatical unit in Māori is a phrase.
As mentioned earlier, Croft’s (1990, 2001, 2004) prototype analysis succeeds in untangling the confusion that always accompanies the discussion of the parts of speech. The English -ing form, for example, has both nominal and verbal properties. *Swimming* as in *Swimming is the only exercise I can do without hurting my knees* is an action-denoting reference expression, and thus finds its place in the lower left box in the map. *Swimming in swimming pool* fits in the lower middle box, as it is an action-denoting modifier. *Swimming in Henry is swimming competitively* is in the lower right box, as this *swimming* is in the paradigm of the verb *swim*. It is marked, despite its full-fledged verb membership, but that is what Croft explains best: Categories are not defined by sharp demarcations that require all the members to show all the necessary properties for the category, but by identifying the core members with the most desired properties, some of which are shared by other members to be included and considered for membership. Therefore, the full continuum of *swim* (v) goes from a prototypical unmarked use (*Dogs can swim*) to the least prototypical marked use (*Swimming is fun*), and somewhere in between, on the continuum, *swimming* as a participle used in the present progressive and for modification can be located. English speakers know by analogy that *swim* in *one-mile swim* is a derived noun, unmarked yet non-prototype; it belongs in the lower left box of the map as a deverbal noun (via zero-derivation). Therefore, Croft’s table is not wrong. The reason for the inconsistent patterns that Māori base items show in Croft’s table is that Māori is an exception.

**5.7. Category-less Bases**

Chapter 4 introduced the category underspecification approach, which is taken by many linguists, especially those who study non-Indo-European languages (e.g., Foley for Philippine languages [1998], Broschart for Tongan [1997], Peterson for Kharia [2013], Gil for Riau...
Indonesian [2000, 2013], Bisang for Late Archaic Chinese [2010, 2013]). The main claim of this approach is that the lexical items are not preclassified for syntactic slots—contrary to the claims of the generative theories. Therefore, for such languages with category-free lexical items, a clear distinction has to be made between the lexical and syntactic categories, and it is the syntactic categories that the grammar is concerned with.

Māori is a strong candidate for such a language, for three reasons: (1) the basic grammatical unit is larger than a word (base); (2) the bases cannot appear in the sentence by themselves (except in a few particular constructions); (3) there are two distinct phrase categories, DET-phr and TAM-phr, to either of which many bases can attach; and (4) bases do not change their forms when used with DET and TAM. First, the phrase being the primary grammatical unit in Māori supports the separation between syntactic and lexical levels, and the syntactic level is of primary importance in Māori grammar. That is, Māori definitely has syntactic categories, which are the TAM-phr and DET-phr categories. Second, the bases’ dependence on TAM and DET to be used in a sentence strongly suggests the possibility that base items by themselves, in isolation, have no designated grammatical function. For example, the base mōhio is ambiguous in terms of its meaning (‘know’ or ‘knowledge’ or ‘knowledgeable’) and function (predicate or reference or modification), when it is not preceded by TAM or DET. Recall that a category label such as “noun” specifies a word’s function before it is inserted in a sentence; a noun prototypically functions as a reference to an object, which in turn means it is used as an argument in a sentence. Thus, when a word is not specified for category, its grammatical role in the sentence is also unspecified. That is, Māori bases in isolation have meaning (or concept, to be precise) only. The concept, however, has the “potential” to appear in various contexts and to be used for certain meanings. Last, when words do not change their forms, the only positive evidence for separate
categories is the existence of other words that do change form according to category. English, for instance, has a paradigm for verb inflections, which serves as a basis for analogy for English speakers to determine that words such as hit and cut are also nouns. Derived nouns such as development and creation can signal their noun membership even in isolation because of the presence of the derivational suffixes -ment and -tion, which is why they cannot be used as the head of VP or as noun modifiers: Developing/*Development countries suffer more from natural disasters, or Positive thinking creates/*creations positive results. These examples enable us to identify cut in We have to make some cuts in the budget and hit in The hurricane made a direct hit as derived nouns. For Māori, the majority of the words do not change form, and there are no other cases that serve as the standard or point of reference to establish the categories. There are only two phrasal categories, in both of which the majority of the base items can appear “as is.” There is no morphological evidence that leads us to category distinction in Māori bases.

When there are two distinct phrase categories and the majority of the bases can appear in either of them without changing their forms, the categories of the bases only need to be restrictive enough to account for these facts. If the bases came with pre-assigned categories, it would have to be because Māori syntax required lexical information, and this would not allow the categories to overlap, because such categories would be useless. These facts point to the necessity of a precategorial approach for Māori. However, Māori does not have completely flexible word class. There are productive suffixes such as -Canga and -Cia that select only one type of preposed periphery particle. There are some words that only take DET. The precategorial approach must be able to explain these facts as well if it is to account for Māori grammar.
5.8. Māori as a Type/Token Language

Among the languages that were introduced in chapter 4, Tongan exhibits similar category overlap phenomena as found in Māori. Because they both belong to the same Polynesian subgroup of Oceanic languages, it is no surprise that they share similarities. Broschart’s (1997) analysis of Tongan may not apply directly to Māori, but it has valuable insights for comprehending Māori.

As briefly explained in chapter 4, Broschart claims that the use of tense markers does not change the meaning of expressions in Tongan (1997:137):

(5.11) ‘oku fu’u fo‘i ‘ulu lanu pulū e kakaā
PRES CL.big CL.round head colour blue DET parrot.DEF
‘The parrot has a big round blue-coloured head.’

(5.12) ko hono o‘i ‘ulú
PRST POSS.IN.3SG CL.round head.DEF
‘its (round) head’

According to Broschart, ‘ulu ‘head’ does not change its meaning whether it occurs with the tense marker ‘oku as in (5.11) or with ko as in (5.12). ‘Ulu is not “verbalized” in (5.11); rather, the TAM context “brings about the different interpretation” (Broschart 1997:137). In the next pair, pa‘anga ‘dollar’ does not become an “action” just because it is preceded by the TAM ‘oku in (5.13).

(5.13) ‘oku pa‘anga ‘e tolu
PRST dollar LNK 3
‘It is three dollars (in value).’ (Broschart 1997:138)

(5.14) Ko e pa‘anga ‘e tolu
PRST DET dollar LNK 3
‘(It is) three dollars (as coins, etc.) (Broschart 1997:138)
The meaning of *pa‘anga* ‘dollar’ remains the same, but the time-stable interpretation of the word *dollar* is brought out in (5.14), as in *three dollars in coins*, with the co-occurrence with ART *e*, while the TAM *oku*, which does not imply time stability, brings out the interpretation that implies a transitional state, such as the value of something.

The English translation gives the impression that the use of TAM with a “noun” converts the noun’s meaning to allow it to fit in verbal and predicate uses, that the meaning after the conversion is idiosyncratic and unpredictable, and that this is the nature of zero-conversion. Broschart (1997:142) argues against this view with the example of *te u puaka* (FUT 1sg pig), which means ‘in the future (there will be) (a non-referential) pig-manifestation (as provision) of me’. Tongan speakers, using their pragmatic and cultural knowledge of pigs being provisions can interpret the sentence as ‘I will provide pig’. In other words, the meaning of the bases that appear with TAM and DET does not change, but their interpretations are determined by the meanings of the lexical items, which restrict the possible interpretations on their own. That is, *puaka* ‘pig’ combined with tense will never have the ‘is worth’ interpretation of *pa‘anga* ‘dollar’, and *ulu* ‘head’ combined with tense will never have a ‘provide’ interpretation. The interrogatives are also a good example for this claim. Broschart includes examples of Tongan interrogatives with TAM:

(5.15) ko e hā?
PRST DET what
‘What is that?’ (Broschart 1997:138)

(5.16) ‘oku hā?
PRES what
‘What was that?’ (Broschart 1997:138)

---

36 This is the translation given by Broschart. Dr. Yuko Otsuka (pers. comm.) points out that the sentence can also mean ‘what is happening?’ or ‘what about it?’ or ‘(he) does what?’ (as an echo question).
The interrogatives do not, or cannot, change meaning as they appear with different particles. The point is that there is no “conversion” or derivation in a certain direction between the lexical items appearing in TAM-phrase and DET-phrase.

Broschart makes the further claims that tense in Tongan “functions as a presentative element instantiating a propositional relation between whatever lexical item and its argument(s) and it is the proposition instantiated (e.g., the relationship between a body part and its possessor) which is ‘situated’ in time.”

(5.17) na’e ‘alu a Sione ki kolo
   PST go ABS Sione to town
   ‘Sione went to town.’ (Broschart 1997:141)

The sentence in (5.17) can be literally translated as ‘in the past was going (of) Sione to town’.

Broschart questions why there are two distinct phrase types in Tongan but the lexical items do not recognize the difference and can select either of them. He proposes that Tongan ART-phrase and TAM-phrase share the feature [+ref(erential)], while the lexical items in isolation are [-ref]. Other languages that have a noun and verb distinction do not have this primary division, as the lexical items are divided into nominal and verbal, which correspond to the defining feature [-pred] and [+pred] respectively. This correlation is observed by morphological or morphosyntactic markings (Hengeveld’s ([1992:58] “further measures”); for instance, nouns need copula for predication, and verbs need to be nominalized for referential use. This, apparently, is not the case for Tongan, which means that [± pred] is not the most decisive criterion for word classes. Instead, Broschart proposes that Tongan lexical items are primarily divided by referentiality: Tongan lexical items are referentially unsaturated ([-ref]), while the phrases, both TAM and ART, are referentially saturated [+ref], because one can relativize both the TAM-phrase and the ART-phrase (e.g., ‘It was raining, which is unfortunate’ and ‘The lion which I
saw yesterday’). Lexical items with the feature [-ref] are “types,” and those with [+ref] are “tokens.” The types are the forms that represent an abstract concept. The type ‘alu ‘go’, for example, represents the concept of GO—something about going—that has several possible projections and opportunities in proposition. However, ‘ALU as a type needs to become a [+ref] token, which is the functioning grammatical unit, by combining with TAM or ART to be inserted in syntax. Therefore, tokens are syntactic units, while types are lexical items without categories. Whether or not the expression is more suitable for predication is relevant to the items on the phrase level, not to the items on the lexical level.

Because a type contains only concept, no other information about the item, including the category, exists. Tokens, on the other hand, fall into the (syntactic) categories of ART-syntagm and TAM-syntagm, which are syntactic (phrasal) categories. Thus, Tongan does not have lexical categories inherent to lexical items, nor noun/verb/adjective distinctions on the lexical level. Broschart claims that, without the noun/verb/adjective distinction, there cannot be nominal/verbal/adjectival phrases, while the ART-syntagm can be equated to DP, and the TAM-
Many of Broschart’s insights reviewed in this discussion apply to Māori. First of all, co-appearance with TAM does not “verbalize” a base item. Just as Broschart explained for the Tongan word *pa’anga* ‘dollar’, it does not become a “verb” and acquire a different meaning ‘is worth’ when it is attached to TAM. In Māori, too, a TAM marker does not make bases into verbs.

(5.18) i a ia e tamariki ana
at PERS 3SG TAM children TAM
‘when he/she was young’

*Tamariki* ‘children’ in (5.18) is not in a category “verb” simply because it co-occurs with TAM. Rather, the progressive marker *e…ana* brings out the most appropriate meaning that suits the TAM context from the possible semantic readings of the type, or concept, *TAMARIKI*. For *TAMARIKI*, the transitional, non-time-stable interpretation is ‘being a child as a phase of one’s life’ or ‘being in a childhood stage’. Therefore, *e tamariki ana* can be translated as ‘being young’. The combination of the particle and the base item thus creates the meaning in the given context, understood and accepted as the correct and meaningful use.

However, and also as in Tongan, it is not always the case that the lexical item co-occurring with TAM gets the verbal interpretation and the lexical item co-occurring with DET gets the nominal interpretation. In example (5.19), two so-called “verbs” are in a coordination construction:

(5.19) Ka umere me *te* kata
TAM shout and DET laugh
‘(They) shouted and laughed.’

Despite the fact that the two actions are performed by the same subject, the second action is in a DET-phrase. There is no explanation for the different selections of preposed particles, except...
that Māori coordination requires this construction. This illustrates that the TAM-phr is not necessarily a marker for a verbal meaning, nor the DET-phr for a nominal meaning.

Indeed, Māori makes use of DET-phrases in constructions in which verbs would be used in their English counterparts.

(5.20) He aha te mahi o Hone?
DET what DET work of Hone
‘What does John do for work?’ (Lit: What is the work of John?)

The expression of exclamation is simply a te-phrase:

(5.21) Te tere o te oma nga mai
DET fast of DET run.NMLZ hither
‘How fast he runs!’

The te-phrase can be translated as a sentence in English. Bauer calls this “independent assertion”: 

(5.22) Te puta ana mahara i te kore waka
DET pass her thought at DET NEG canoe
‘Her thoughts turned back to her lack of canoe.’ (Bauer 1997:528)

Note that, as in (5.22), te can also appear with the negative expression kore.

Te + -Canga is used frequently in the subordinate clause. The following example (5.23) shows that in a coordination construction, the first DET-phr is the subordinate clause, marking the event that happens prior to the event in the main clause.

(5.23) Te mutu nga o te haka o Pipito, ko te puta nga o Wairangi
DET end.NMLZ of DET haka of Pipito, EQ DET appear.NMLZ of Wairangi
‘When Pipito’s haka ended, then Wairangi appeared.’ (Bauer 1997:520)

It appears in a reason clause:

(5.24) Nā te haere nga mai o Paki, i oti ai te mahi of DET go.NMLZ hither of Paki, TAM completed ANAPH DET work
‘Because Paki came, the work was completed.’ (Harlow 2001:214)

And it appears in a conditional clause:
(5.25) Ki te ua āpōpō kāore au e haere
P DET rain tomorrow NEG 1SG TAM go
‘If it rains, I will not go.’

As shown by the English glosses, Māori makes use of DET-phrase constructions frequently, where verbal constructions are used in English. Therefore, the forms that are “nominalized” by -Canga or the forms derived by “zero-derivation” are used in the te-phrase constructions. Bauer (1993: 254-255) warns that this determiner te should not be the criterion for nounhood as it is too inclusive. The determiner te also appears as a part of some TAMs, such as the progressive kei te and i te, and also in the infinitive clause for purpose, ki te. The progressive markers kei te and i te are considered lexicalized, with each taken as a unit, and the base that follows ki te is translated as a verb in its infinitival form:

(5.26) Me haere au ki te whare wānanga ki te ako i te reo Māori
must go 1SG to DET university to DET learn DO DET language Māori
‘I must go to the university to learn Māori.’

Harlow (2007) notes that, although ki te for a purpose clause is identical to the preposition and determiner string, ki te, in ki te whare wānanga ‘to the university’, the purpose clause ki te string is now considered a TAM in modern Māori. If the purpose clause ki te is a TAM, then ki te used for the conditional clause (5.25) should also be taken as a TAM. These examples show the influence of English in the analysis of Māori grammar. Because the English translations in (5.25) and (5.26) use ‘if it rains’ and ‘to learn Māori’, the Māori equivalents are also treated as verbs, and the preposition-determiner string ki te is reanalyzed as a one unit TAM. The progressive markers kei te and i te must have each become a single unit in the same way in the past and are now used as TAMs. They are treated as TAMs in all three major reference grammars of Māori (Biggs 1969, Bauer 1997, and Harlow 2001), and as verb markers in Māori textbooks. Although such reanalysis is plausible for the infinitival ki te, there is no strong evidence for it, except that
the relevant constructions are translated as verbal constructions in English. As native Māori speakers are bilingual English speakers, the English influence is inevitable. However, is it necessary to change the categories to align with the translation?

It is equally plausible that \textit{kei te} and \textit{i te} still are prep + DET. The following sentence is ambiguous:

(5.27) Kei te mahi a Mere
\hspace{1cm} P \hspace{0.5cm} DET work PERS Mere
\hspace{1cm} ‘Mary is at work.’

(5.28) Kei te mahi a Mere
\hspace{1cm} TAM \hspace{0.5cm} work PERS Mere
\hspace{1cm} ‘Mary is working.’

Harlow (2001) writes that “the main function of the preposition \textit{kei} is to introduce predicate phrases stating where the subject is located” (80) and goes on to explain that \textit{kei}, along with \textit{i} (for past) and \textit{hei} (for future) “assert that the subject is located in the place mentioned in the predicate phrase” (2001:140). The construction Harlow refers to is given in (5.29).

(5.29) Kei te whare a Hōne
\hspace{1cm} P \hspace{0.5cm} DET house PERS Hone
\hspace{1cm} ‘John is at home.’

This construction is clearly identical to those in (5.27) and (5.28), except that what follows \textit{kei} is a place. The only reason that one feels the urge to posit a different construction for (5.28) is the English translation. However, (5.28) can be interpreted as ‘Mary is in the state of working (in the present time)’, which is, in correct English expression, ‘Mary is working’.

(5.30) Kei te haere a Mere ki te kula
\hspace{1cm} P \hspace{0.5cm} DET go PERS Mere to DET school
\hspace{1cm} ‘Mary is going to school.’

The example means, ‘Mary is in the state of going to school (in the present time)’.
(5.31) I te kai ngā tamariki
    P DET eat DET children
    ‘The children were eating.’

This can be expressed as ‘the children were in the state of eating in the past’. It definitely sounds awkward in English, but this is because English does not express the same meaning that way. *Kei te* analyzed as TAM has its purpose (i.e., for ease of understanding for Māori learners, for instance), but the original meaning does not have to be thrown away.

For *ki te*, as well, it is unnecessary to discard the original prep + DET label. Harlow (2007:191) notes that *ki te* in the sentence such as (5.32) is analyzed as TAM in modern Māori:

(5.32) Ka haere a Mere ki Pōneke *ki te* kanikani
    DET go PERS Mary to Wellington to DET dance
    Mary is going to Wellington to dance.’

The literal translation is ‘Mary is going to Wellington to the dance/dancing’. The preposition *ki* has many interpretations. It can mark the direct object for experience verbs and it means ‘towards’ with motion words, ‘to’ for indirect object, ‘until’ with temporal expressions, ‘with’ for instruments, and so forth. While the preposition *ki* does have many functions, some of these interpretations are also due to the translations. *Ki* is flexible in its interpretation in the sense that it connects events (e.g., Mary’s going) and places (e.g., Wellington) and reasons (e.g., *kanikani*). Therefore, instead of reanalyzing *ki te* as the TAM for an infinitival clause to match the English translation, giving *ki* an additional function “for the purpose of” is another option. Thus, *ki te kanikani* means ‘for the purpose of dancing’, which, in more appropriate English, is ‘in order to dance’. It is not only unnecessary but also futile to accommodate Māori grammar to English.

(5.33) He mōhio a Moana *ki te* waiata
    DET knowledgeable DET Moana to DET song
    ‘Moana is knowledgeable about song.’ (Waite 1994:59)
For the sentence in example (5.33), Waite glosses *mōhio* as ‘knowledgeable’, and *he* as TAM, which takes an AP complement, in spite of the fact that *he* is an indefinite determiner, often introduced in Māori textbooks with the explanation: “*he* is indefinite, *te* is definite.” However, *he* and *te* have different distributions: *he*, unlike *te*, cannot follow prepositions. This means that *he*-marked phrases cannot be used as arguments except for subjects, which are unmarked (i.e., without prepositions). *He*-marked phrases can be predicates, as in *he tākuta ahau* ‘I am a doctor’, without any predicate marker or copula. Because Māori is a VSO language, the sentence-initial position is recognized as a predicate position. Therefore, *he* in the sentence-initial position is analyzed as TAM. However, this is not necessary, if one prefers to keep a more consistent analysis of *he* as a determiner. That is, *a Moana* can be analyzed as the subject (i.e., with *a* as personal article), but if *a* is analyzed as the possessive particle, then *he mōhio a Moana ki te waiata* ‘Moana’s knowing of the song’ as a whole is the predicate of the sentence. This means that this sentence does not contain a subject, which occurs in Māori, as seen in (5.21). In the subject-less construction, the notional subject is often introduced as a possessor, as we will see in the example (5.54). The advantage of this interpretation of (5.33) is that we do not have to posit two different functions for *he* (one as DET, another as TAM), and consequently *mōhio* ‘know’ does not have to be interpreted as an adjective ‘knowledgeable’. The determiner status does not affect the base *mōhio* in the precategorial approach, because *he mōhio* (‘Det know’) can be interpreted in the context appropriately as ‘Moana is the know’, meaning ‘Moana is a knowledged one’. The most natural English expression may use ‘knowledgeable’ instead of ‘a knowledged one’, but that is not a Māori issue, it is an English issue. Māori grammar can and should keep its own linguistic integrity.

In Māori, the bases that appear with TAM are all able to appear with DET, which Bauer
(1997) calls “stem nominalization.” Stem nominalization includes any “verbs” appearing with DET: the purpose clause ki te, the construction for exclamation te + ‘adjective’ (as in te tere ‘how fast’), the te-marked “verbs” used in the argument positions, and so forth. As stem nominalization is used frequently, Bauer warns that two identical-looking constructions may be the same in appearance only. According to her, the following two identical sentences are actually different constructions, as suggested by the translation:

(5.34)  Ka hiahia ia ki te hoe
        TAM want 3SG to DET paddle
        ‘He wants to paddle.’ (Bauer 1997:528)

(5.35)  Ka hiahia ia ki te hoe
        TAM want 3SG to DET paddle
        ‘He wants the paddle.’ (Bauer 1997:528)

Bauer considers the first te hoe to be verbal, because one may say:

(5.36)  Ka hiahia ia ki te hoe i te waka
        TAM want 3SG to DET paddle DO DET canoe
        ‘He wants to paddle the canoe.’ (Bauer 1997:528)

The fact that the direct object follows it suggests that te hoe in (5.36) is a verb introduced by DET, whereas for (5.37), the genitive marked phrase follows:

(5.37)  Ka hiahia ia ki te hoe o te waka
        TAM want 3SG to DET paddle of DET canoe
        ‘He wants the paddle of the canoe.’ (Bauer 1997:528)

Te hoe in (5.35) and (5.37), therefore, is a noun, derived from the verb hoe. Recall that Waite (1994:60) gave the same analysis for the examples repeated here:

(5.38)  Ka pakaru te wini i te [v waiata] a te wahine
        TAM broken DET window DO DET sing of DET woman
        ‘The woman’s singing broke the window.’

37 Stem nominalization is also possible with he and hei.
If Broschart’s approach is applied to these examples, they can be analyzed as identical constructions. Examples (5.34) and (5.35) are ambiguous. Whether te hoe means the action of paddling or the paddle as an instrument for paddling is unknown without a bigger context. If (5.34) is the answer to the question of what somebody’s son wants to do when he goes to college, te hoe means the action of paddling, and if (5.35) is what the son wants for his birthday present, te hoe means the instrument of paddling. Without such information, (5.34) and (5.35) are the same construction, not two different constructions. Furthermore, what type of phrases can follow is determined by the meaning of the base, not by some kind of hidden or underlying construction that somehow signals whether to take the DO phrase or the genitive phrase. It is not that te hoe in (5.36) is a verb because it is followed by a DO phrase; it is followed by a DO phrase because te hoe in (5.36) is interpreted as the action of hoe in the given context. Note that, contrary to Bauer’s (1997) claim, Waite (1994) labels te waiata in (5.38) a verb, even though it is followed by the genitive a. Thus, in short, if one can stipulate that “determiner te can also take a verb,” one can also assume that there are no nouns and verbs. They are types HOE and WAIATA, and they combine with either TAM or DET. This eliminates the unnecessary invisible derivational process, “nominalization,” which nominalizes even the lexical items that are nouns, and it also eliminates the need to posit two types of nominalization—one keeping the original category and another changing the category—in spite of the identical word forms.

As even Biggs (1969) acknowledged by keeping it, the class of nouns is difficult to reject. While all the bases that take TAM can also take DET, there are base items that co-occur only with DET.
Bases such as *ika ‘fish’ are prototypical nouns according to prototype analysis (Croft 2001, 2004), as they denote objects while their function is reference. It is also intuitively very convincing that all languages have nouns at least. It appears that Māori has nouns, which are those items that can only be preceded by DET.

Māori has phrases such as *enga ru ana te moana ‘the sea is rough today (lit. the sea is wave-ing)’, *ka tangata te tamaiti ‘the child has become a man’, and *ka rākau te taone ‘the town is covered with trees (lit. the town has tree-ed)’. In these constructions, *nga ru ‘wave’, *tangata ‘man’, and *rākau ‘tree’ take TAM, and the English translation needs various additions to make sense, such as ‘become a man’ and ‘is covered by trees’. Broschart claims that there are limited possible interpretations for certain words when they are combined with TAM or DET. In Māori, when a word denoting humans such as *tangata attaches to TAM, there is only a limited number of possible interpretations, one of which is ‘to become a man’, most likely due to previous usage. The usage and the meaning work with other similar lexical items referring to human beings such as *koroua ‘old man’ and this particular interpretation becomes a pattern. *Kua rākau te taone means ‘the town is covered by trees’, whereas *ka rākau can also mean ‘has become a tree’ for some speakers. The lexical items have patterns for possible interpretations when they are used in DET-phr or TAM-phr. Bisang (2013:286) finds recurring patterns in interpretation when object-denoting lexemes appear in the verbal slot in Late Archaic Chinese. Giving the examples *chéng (N) ‘wall’: *chéng (V) ‘to wall a city’, and *jūn (N) ‘army’: *jūn (V) to
encamp’, Bisang suggests that if precategoriality is a basic property of lexical items, and if their assignment to the verbal or nominal slots is governed by pragmatics, it is to be expected that certain object-denoting lexemes get specific interpretations if they occur in recurrent situations. Clark and Clark (1978) claim that denominal verbs require the specific time, place, and circumstances in which they are uttered in order to be used successfully in the speech community. As Clark and Clark note,

the speaker means to denote the kind of state, event, or process that, he has good reason to believe, the listener can readily and uniquely compute on this occasion, on the basis of their mutual knowledge, in such a way that the parent noun (e.g. porch or Houdini) denotes one role in the state, event, or process, and the remaining surface arguments of the denominal verb denote others of its roles.

(Clark and Clark 1978:767)

What Clark and Clark refer to as denominal verbs are not the verbs regularly derived from nouns such as bicycle or smoke (their examples), but porch or Houdini, that is, newly created innovations. Although they claim that such innovations differ from regular derived verbs in the sense that they require specific circumstances for their creation and acceptance, they also suggest that many well-established denominal verbs were once such innovations. Following this line of thought, in Māori, while kua tangata ‘has become a man’ is a well-established denominal verb, some speakers do not find kua rākau ‘is covered by trees’ acceptable, perhaps because its denominal process may still be in progress. However, this is not reason to deny the fact that it has the potential to appear with
TAM if the need arises. And what this suggests is that, although there are some “prototypical nouns” that never appear with TAM, which is the reason that Biggs kept the class of nouns, they all can potentially appear with TAM in appropriate contexts. That is, kōhatu ‘stone’, which can take TAM to mean someone turned into a stone, is not a different type of word from putiputi ‘flower’ because the former is compatible with TAM (thus, a “universal”) and the latter is not compatible with TAM (thus, a “noun”). That kōhatu has TAM + form is “accidental”; there was once a need to use kōhatu in a TAM context, in a legend for example, and DET + kōhatu was an innovation. An English denominal verb such as xerox in *I need to xerox some more for tomorrow’s meeting* is accepted, while other brand names of copy machines are not acceptable in the same context. The difference is not grammatically motivated: It is outside the scope of grammar. Therefore, although there are bases that are not used with TAM, that does not mean they constitute a separate category from bases that are compatible with TAM. All Māori bases are potentially TAM and DET friendly.

5.9. Māori Phrase Structure Revisited

We have entertained the possibility that the base items in the nuclei in Māori are category-free. This approach frees us from the task of finding ways to explain the inconsistencies that we face if the noun/verb/adjective distinction is forced onto Māori. The bases in Māori are types that carry the general concept, until they become tokens, fully-functioning lexical items in syntax, and all the bases have the potential to appear in either a TAM-phrase or a DET-phrase. TAM and DET do not form separate paradigms like nouns and verbs, but they belong to the
same syntactic slot within the contour phrase. The difference between TAM-phr and DET-phr is that TAM-phr cannot follow the preposition. Bases in nuclei can choose ONE particle from the preposed periphery, while the particles in the postposed periphery are optional. Table 5.8 below represents the contour phrase’s internal structure.

**Table 5.8. Māori contour phrase internal structure**

<table>
<thead>
<tr>
<th>TOKEN TYPE</th>
<th>PREPOSED PERIPHERY (particles)</th>
<th>NUCLEUS (bases)</th>
<th>POSTPOSED PERIPHERY (particles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAM</td>
<td>ka, kua, i, kei te, i te, e (etc.)</td>
<td></td>
<td>mai, atu, tonu, rawa, nei, ra, ai, ana</td>
</tr>
<tr>
<td>DET</td>
<td>te/ngā, he, tētahi/ētahi, hei (etc.)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(lexical items that can appear without TAM/DET) *kaore, ehara, katoa, au, koe, ia, etc.*

Table 5.8 should not mislead readers to the understanding that any combination of particles and bases is allowed. There are constraints semantically, discoursally, and pragmatically. This table illustrates the tripartite structure of a Māori phrase, in which the bases combine with either TAM or DET as long as they do not violate non-syntactic requirements.
In English, for example, the nouns and verbs form separate paradigms because of their different inflectional patterns. In Māori, as seen in the table above, TAM and DET do not form separate paradigms and they share the same slot in the contour phrase’s internal structure. This is a telling fact about TAM and DET: They form two distinct types of phrases (*preposition + TAM-phr, preposition + DET-phr), yet they are the same in distribution and in compatibility with the bases that follow. Distinctions or relationships between TAM and DET appear to be fundamentally different from distinctions or relationships between “verbal” and “nominal.” It is possible that Māori TAM and DET form a continuum. If this analysis is correct, it explains TAM ka, which is tense-less, aspect-less, and mood-less, yet functions as a TAM. Its function is to create a predicate:

(5.42) Ka haere a Hone
      TAM go PERS Hone
      ‘Hone goes.’

If the temporal adverb āpōpō were present in the sentence in (5.42), it would give a future reading. Ka is used in the main, not the subordinate, clause, and the ka-marked phrase can be used in a narrative once the time is set by other TAMs. However, compare the two sentences below:

(5.43) Ka nui te whare
      TAM big DET house
      ‘The house is big.’

(5.44) He nui te whare
      DET big DET house
      ‘The house is big.’

Both sentences are describing a property of the subject whare ‘house’. There is certainly a clear difference between ka and he, because the following examples are ungrammatical:
(5.45) *Ka ika ahau
  TAM fish 1SG
  ‘I fish.’

(5.46) *He pānui ia i te pukapuka
  DET read 3SG DO DET book
  ‘He read the book.’

Ka is definitely a TAM, which describes nui as a temporal state, while he as a determiner presents nui as a set property of the house. What I propose here is that the TAM ka is the most DET-like of all the TAMs. Bauer shows this example:

(5.47) Ka nui te mīti, te hēki me te tuna
  TAM big DET meat DET egg with DET eel
  ‘There’s plenty of meat, eggs and eels.’ (Bauer 1993:87)

The English translation does not mean that the TAM ka has an optional or hidden function as an expletive. To translate the literal meaning of ‘the meat, eggs and eels are big in quantity (at this moment in time)’, the English expression ‘there is’ is the most appropriate. However, the fact remains that the TAM ka has no verb-ness in (5.47); rather, it almost has a presentative quality, much as Broschart (1997) suggests about Tongan tense (see section 5.8).

If this interpretation of TAM is not wrong, it is also possible to introduce an alternative interpretation of the controversial DET he. The peculiar behavior of he is a topic of investigation among Māori scholars. To summarize, he is a non-specific determiner (as in he rākau ‘a tree’), which has a very limited distribution. A he-marked DET-phrase is used only for subjects,38 which means he-marked phrases are not compatible with prepositions.

(5.48) *Kua tuhituhi ahau i he reta ki taku hoa
  TAM write 1SG DO DET letter to my friend
  ‘I’ve written a letter to my friend.’

---

38 Bauer (1993) claimed that he cannot introduce the subject of transitive verbs, but revised this statement in her 1997 publication, where she proposed that he can mark all subjects in certain semantic environments.

39 Te or tētahi could replace he, or the passive construction could be used to make he reta into a superficial subject.
He-marked phrases introduce predicates:

(5.49) He mōmona te kurī
DET fat DET dog
‘The dog is fat.’

(5.50) He kaiako a Mere
DET teacher PERS Mere
‘Mary is a teacher.’

Also, as seen in (5.33) earlier, when the determiner he appears in the predicate, it can be analyzed as TAM (which is glossed as TNS in this example from Bauer and Bauer 2012:10):

(5.51) He uaua te mōhio.tia o ngā whakaaro o ngā kaihaina
DET difficult DET know.PASS of DET thought of DET signatory
‘It is difficult to know what the signatories had in mind.’

The restrictive nature of he-marked phrases as an argument is discussed thoroughly by Chung et al (1995). What is significant for my argument is that he is given a function as TAM. As mentioned numerous times, TAM and DET form two distinctive phrases in Māori. In such a language, it is highly unusual that the same particle is used for both TAM and DET. There are instances in which the same particle functions as TAM and preposition, such as i and kei, or DET and preposition hei, which is puzzling also, but particle overlap for TAM and DET poses a crucial problem in understanding the Māori propositional mechanism. However, if TAM and DET are not two separate entities, but rather possess graded properties, it is possible to understand he as the most TAM-like DET.

(5.52) He reka te wai
DET sweet DET water
‘The water is sweet (tasty).’

In (5.52), literally meaning ‘the water (is) a sweet/sweetness’, it is possible that he, possessing some quality of TAM [+pred], has a functional opportunity to introduce the predicate. This suggestion may be as stipulative as labeling it TAM for convenience. However, in a type/token
language, \([\pm \text{pred}]\) is not as salient a feature as it is for languages that have the nominal/verbal distinction. There are many examples of DET-phrases used for verbal interpretations, as in (5.53). Labeled an “existential” sentence, one he-marked phrase can make a complete proposition:

(5.53) \textbf{He taniwha}  
\textbf{DET taniwha}  
‘There are taniwhas/Taniwhas exist.’ (Bauer 1997:34)

Bauer explains that it is a predicate, as he indicates a class of \textit{taniwha}, not one specific \textit{taniwha}, and this construction is used for existential sentences. Unlike English, which requires an expletive or the verb ‘exist’, the literal translation of this example is ‘A taniwha’. \textit{He} must possess a quality in this construction that must signal its predicate status. Harlow gives the following example for the use of \textit{he}:

(5.54) \textbf{He moni ā rātou?}  
\textbf{DET money their}  
‘Have they any money?’ (Harlow 2007:165)

According to Harlow, ā \textit{rātou} ‘their’ is a subject and \textit{he} \textit{moni} is the nominal predicate. Languages that do not have the verb ‘have’ express possession with existential verbs in constructions such as ‘the money exists to me’. Therefore, \textit{he} in (5.54) may also imply existence, as in (5.53). These examples show that we cannot characterize TAM for predicate and DET for non-predicate in Māori. Nominal predicates are often used in Māori, and a \textit{he}-marked predicate is as “marked” as a TAM-phrase predicate, in the sense that both require preposed particles. The functional differences between Māori TAM and DET are fuzzy. While \([+\text{pred}]\) is clearly the property that goes with TAM, \([-\text{pred}]\) for DET should be taken only as a primary criterion to characterize DET, because, unlike TAM, which can never be used to introduce the argument, DET can be used as the predicate. The only and decisive criterion that separates TAM and DET
is the ability to follow prepositions, which means that a phrase either can or cannot be used as an argument: TAM-phr cannot, while DET-phr can. Figure 5.5 summarizes the Māori base classification.

![Figure 5.5. Classification of Māori base items](image)

**5.10. Affixes as Type**

An assumption that base items in Māori are category-free brings numerous benefits. It simplifies the account of Māori grammar, because it eliminates zero-derivation and the mismatch between the lexical and phrasal categories, and it enables us to make a uniform analysis of the tripartite phrase-internal structure that sheds light on the unique nature of Māori TAM and DET. With this approach, it is never necessary to change the grammatical functions of particles, such as by reanalysis of the purpose clause marker *ki te* to TAM. However, there is one more issue: affixes. In this section, I discuss how the affixes can be accounted for in the category
underspecification analysis, using as examples the two most productive suffixes, -Canga and -Cia, which were presented in chapter 4.

Recall that Biggs (1961) refers to these suffixes as “nuclear minor morphemes,” as they are attached to the bases in the nuclei. Neither the so-called nominalizing suffix -Canga nor the so-called passive suffix -Cia attach to any base indiscriminately, and the suffixed bases can no longer select TAM or DET freely. As their names suggest, “nominalized” bases always choose DET and “passivized” bases choose TAM, and then can be “nominalized.” This causes a problem for an analysis without lexical categories, because the affixes could not choose the stems to which they will attach if there were no category information. However, there is a way around this problem.

I briefly introduced in chapter 4 Lieber’s (2004, 2006) proposal for an affixation process that does not refer to N, V, and A in the stem. In her theory, the lexical item has semantic information (such as, [±animate] [±dynamic], etc.), which also includes relevant syntactic information such as how many arguments the word takes. An affix also has relevant syntactic and semantic information, which should match with that in the stem that it attaches to. Therefore, the affixation process can proceed without referring to the lexical category of the stem. As illustrated in chapter 4, this approach has the advantage of explaining how the agent suffix -er in English can take not only verbs but also nouns and adjectives, by using semantic features for coindexation. For example, write has the semantic feature [+dynamic] and also the information that it subcategorizes two arguments. The suffix -er also has the feature [+dynamic] and it has one argument. The semantic features for both items match, and the affix’s only argument “R” (for “referential”; the external argument of a noun) coindexes the closest argument in the stem.
According to Lieber (2004:16), “the ‘R’ argument of a noun may be discharged by linking it with an NP of which it is predicated (Williams 1981) or by linking with a determiner (Higginbotham 1985; Sproat 1985).” Thus, *writer* gains the agent interpretation.

(5.55) $\{+\text{material}, \text{dynamic} ([i], \ [+\text{dynamic} ([i], [ ])])\} \ 
-er \ 
write

Lieber’s insight is useful in accounting for *-Canga* and *-Cia*, as both of them can attach to multiple categories (that is, if one assumes Māori has categories). In chapter 3, I showed that *-Canga* can attach not only to so-called verbs (e.g., *patu: patu.nga* ‘hit: hit/hitting’), but also to adjectives (e.g., *pai: pai.nga* ‘good: goodness/well-being) and nouns (e.g., *rangatira: rangatira.tanga* ‘chief: chiefliness’). *-Cia* can also take for stems both transitive and intransitive verbs (e.g., *hoko: hoko.na* ‘buy: bought’, *haere: haere.a* ‘go: (is) gone over’), as well as nouns (e.g., *rākau: rākau.tia* ‘tree: tree-ed’). Both suffixes can attach to interrogatives (e.g., *aha: aha.tanga* and *aha.tia* ‘what: what thing’ and ‘what-ed’), and they have to attach to the modifier as agreement (*tae.nga tuatahi.tanga* ‘first arrival’, *whakaputa.ina tuatahi.tia* ‘first published’).

In “nominalization,” *-Canga*, as well as DET *te*, can attach to virtually any base item to be used in DET-phrases, while *-Cia* has constraints: It cannot attach to some bases such as *pakaru* ‘be broken’, *pau* ‘be completed’, *pai* ‘good’. *Pai* does have a passive form *pai.ngia*, which is used when *pai* means ‘to like’.

For something to be called nominalization, it is reasonable to assume that the derived

---

40 The “R” itself does not have any thematic role, which means the thematic role of the base that “R” coindexes will be absorbed.
base needs to be a verb and needs to be derived from a verb. Māori nominalization and passivization do not show clear patterns. If Lieber’s semantic approach is applied, we can first eliminate the confusing state of base selection, and -Canga and -Cia can be explained more in semantic terms, which might sort the information in a more coherent manner.

First of all, I propose that these suffixes attach to types, rather than tokens. As types, bases such as PATU ‘HIT’ and RĀKAU ‘TREE’ exist as abstract concepts of what the word signifies; the affixes such as -CANGA and -CIA exist as types as well. As Lieber suggests, the semantic features of the bases and affixes need to match (i.e., coindex) for a successful derivation. This derivational process occurs to a type to create another type: PATU ‘HIT’: PATU.NGA ‘HIT-Canga’ and PATU.A ‘HIT-Cia’. We will examine each suffix to understand its semantic function in order to account for the derivation.

5.10.1. -Canga as [-pred]

According to Harlow (2007:122), -Canga derived bases have multiple meanings:

1. an occurrence of V: nō tō rātou hoki nga mai ‘on their return’
2. the product of V: tuhi nga ‘a writing’
3. the place where V takes place: moe nga ‘a bed’

Stem nominalization and -Canga nominalization are similar in the sense that they both create bases that can be used in the argument position or with prepositions. Harlow (2007:122) points out that -Canga derived bases refer to a single event (5.56), just as stem nominalizations refer to a general activity (5.57).

(5.56) nō tō rātou tae nga atu on POSS their arrive NMLZ away ‘On their return’ or ‘When they returned’
Some -Canga suffixed bases have acquired a conventionalized meaning to refer to specific objects, such as moe.nga ‘bed/marriage’ or tuhi.nga ‘writing/essay’. These are the ultimate [-pred] instances. The -Canga form is often found in subordinate clauses for time (5.23) and for reason (5.24), because these subordinate clauses, unlike the matrix clause that offers the main proposition, provide a specific context (or setting) for the event described in the main clause. For instance, in *When the work ends, I will call*, the *when* clause is the “setting” or “specific situation” in which *calling* occurs. Compare that with English, where the nominal phrase is also available: *At the end of the work, I will call*. Aside from the fact that the nominal clause sounds somewhat more formal in English, the nominal usage describes the event ‘the work ends’ more as the point of reference that someone can specifically refer to, and regards the event less as the action of working. That is probably the basic function of -Canga: to lock the base in with [-pred], making the event into SUBSTANCES/THINGS/ESSENCES, to use Lieber’s terminology. Therefore, the feature notation for -Canga looks something like (5.58).

\[(5.58) \quad \text{-Canga} \quad [-\text{pred}, ([], <\text{base}>)]\]

I assume that, like the English suffix -er that has an “R” argument, -Canga has one external argument [ ] in its semantic structure. -CANGA that exists as a type combines with the base and gives rise to a new type, which is still just a concept:

\[(5.59) \quad \text{haere.nga} \quad \text{‘going’ ‘journey’} \quad [-\text{Canga} \quad \text{haere} \quad [-\text{pred}, ([i], [+\text{dynamic}, ([i])] ] ] ] \]
In English, the “R” argument of the affix -er is coindexed with the highest argument of the base to receive the thematic reading and is discharged when it combines with the determiner. In Māori, however, “R” does not necessarily mean “noun,” because both DET-phr and TAM-phr are [+ref]. Therefore, it is the feature [-pred] that selects DET when it becomes a token. The “R” argument that is coindexed with the only argument haere ‘go’ (which is the “subject”) has wide possibilities for interpretation: It describes something to do with going that a person does, as long as it is [-pred]. The newly created -Canga type enters the nucleus as a base that needs to select the particle that has the feature [-pred]; that is, DET. Thus, te haere.nga can be interpreted as ‘going’ or ‘journey’, depending upon the context.

However, there is an interesting example, shown below:

(5.60) …he ara.nga nōna, he whakatika.nga anō hoki tō tana mōkai
   DET rise.NMLZ his DET rise.NMLZ again EMPH POSS his slave
   ‘…he got up, and his slave got up too’ (Bauer 1997:521)

The determiner he, as I have suggested above, is a very TAM-like determiner. Therefore, it is contradictory that -Canga [-pred] is necessary in (5.60), especially as the verbal reading is given in the translation. However, as explained in the discussion of example (5.33), Māori has a subject-less construction; that is, the predicate-only construction functions as a complete sentence. In (5.60) the notional subject of he ara.nga ‘rising’ is marked with the possessive nōna ‘his’, and the notional subject of he whakatika.nga ‘rising’ is marked by tō tana mōkai ‘of his slave’. Therefore, the equivalent English translation is the complete sentence, ‘he got up and the slave also got up’. The type -CANGA attaches to the types ARA ‘RISE’ and WHAKATIKA ‘RISE’ to form new types, which can combine with DET he (as they refer to the specific event) to become tokens.

5.10.2. -Cia as [+high transitivity][+perfective][+affectedness of DO]
Any discussion regarding passives in Māori is complex. To account for the peculiar phenomena in Māori passive sentences, such as their high frequency and their usage in unexpected expressions (in imperatives and with intransitive verbs\(^1\)), the alternative approach suggested in this dissertation, that is, to account for them in semantic terms, needs to look at factors such as the affectedness of DO (Chung 1978: 80), perfectiveness (Clark 1973: 579), and high transitivity (Hopper and Thompson 1980: 252).\(^2\) Bauer (1997) concludes that there is no one factor that explains all the phenomena concerning -\(Cia\). Although a conclusive analysis is beyond the scope of this study, it is still worth entertaining the possibility of analyzing the passive suffix -\(Cia\) in terms of semantic features. First, as with -\(Canga\), let us assume -\(Cia\) attaches to a type, for instance, \(KŌHATU\) ‘STONE’. The type -\(CIA\) would have features such as [+affectedness] [+high transitivity] and [+perfectiveness], instead of “passive.”

\(5.61\) -\(Cia\)

\([ \text{+affectedness/high transitivity/perfectiveness}, ([\text{ }, \text{<base>})])\]

The type -\(CIA\) attaches to a type that is semantically compatible with its features.

\(5.62\) \(inu.mia\) ‘drink-\(Cia\)’

\(-\text{Cia}\)

\([ \text{+affectedness/high transitivity/perfectiveness}, ([\text{i}], \text{[+dynamic ([][-][i])]}))\]

Note that the first argument in the semantic skeleton of the base \(inu\) is not available.\(^3\) This forces the affix to coindex the next argument available. Therefore, the only argument, the patient, receives the semantic features [+affectedness/high transitivity/perfectiveness], and these features

\(^1\) For example, with an intransitive in \(Kua\ oma.kia tōku oma.nga\) (TAM run.pass my run.NMLZ) ‘I have finished my course’ (Harlow 2007:106), and in the imperative \(Inu.mia\ te miraka\) (drink.pass DET milk) ‘Drink milk!’


\(^3\) Lieber (2004:57) is not certain how to analyze passivization in her lexical semantic framework. She suggests tentatively that the passive formation somehow eliminates the highest argument of the base, and the passive suffix creates stative items.
are compatible with the [+dynamic] feature that \textit{INU} ‘drink’ carries. Therefore, the new type \textit{INU.MIA} means ‘something that \textit{INU} affects (highly or completely)’. The \textit{-Cia} derived words select TAM, but not exclusively, as they can appear in a DET-phr:

\begin{align*}
\text{(5.63)} & \quad \text{Mā} & \quad \text{te kōrero.tia} & \quad \text{o te reo Māori e ngā tāngata matatau, ka through DET talk.PASS of DET language Māori by DET people expert TAM} \\
\quad & \quad \text{mau i ngā tamariki. fixed by DET children} \\
\quad & \quad \text{‘It is through the Māori language’s being spoken by those who know it, that it will be grasped by the children.’ (Harlow 2001:191)} \\
\end{align*}

Although this analysis is not confirmed by Lieber, the semantic approach fits well to account for the Māori “passive.” When a type such as \textit{KŌHATU} ‘STONE’ or \textit{RĀKAU} ‘TREE’ is “passivized,” the semantic approach explains the process much more easily, because it does not require a derivational process; that is, \textit{KŌHATU} and \textit{RĀKAU} do not have to become verbs first to be passivized.

\begin{align*}
\text{(5.64)} & \quad \text{kōhatu.tia ‘stone-Cia’} \\
\quad & \quad \text{-Cia} \quad \text{kōhatu} \\
\quad & \quad [+\text{affectedness/high transitivity/perfectiveness ([i], [+material -dynamic, ([i])])}] \\
\end{align*}

The type \textit{KŌHATU} possesses the concept ‘something about a stone/stones’, and it has an “R” argument, which will coindex with the only argument \textit{-Cia} has. Because the head of this operation is the affix, not the base, the non-head base will inherit the head’s semantic features. Thus, \textit{KŌHATU.TIA} means ‘something about \textit{KŌHATU} that affects (highly or completely)’.

How this concept emerges once the item is combined with DET and becomes a token in an acceptable Māori expression depends upon the situation and context.

Taking \textit{-Cia} as an affix with the semantic features [+affectedness/high transitivity/perfectiveness] rather than the passive suffix also explains why some bases, such as
pakaru ‘broken’ and pau ‘used up’ do not take the -Cia suffix. These bases (often labeled “stative” or “neuter” verbs) often denote the state that is the result of the prior action or event. In the next example, for instance, the children’s action resulted in the window being broken:

(5.65) I pakaru te wini i ngā tamariki
   TAM  broken DET  window  P DET  children
   ‘The window is broken by the children/The children broke the window.’

In other words, types such as PAKARU ‘BROKEN’ cannot take -CIA because they denote the state that is already affected by the completed action or event, which is incompatible with the semantic features -CIA carries. Although this analysis is inconclusive and requires a more detailed discussion using more clearly spelled-out features, it suggests the strong possibility that the suffixes -Cia and -Canga are derivational suffixes that create new types. The newly created types select TAM or DET to become tokens in syntactic derivations.

5.10.3. Double suffixation

I have presented a possible account for a suffixation process of -Canga and -Cia that does not require category information, by assigning semantic features to suffixes that seek semantically compatible bases. I proposed that the suffixation occurs to the types, creating new types before they become tokens. The Māori reference grammars mention that the two suffixes usually do not appear simultaneously (Bauer 1997: 516, Harlow 2001: 208).

(5.66) a. Passive
   Ka  patu.a  te  mango  e  te  tamaiti.
   TAM  kill.PASS  DET  shark  by  DET  child
   ‘The shark was killed by the child.’

---

44 Meanwhile, the causative prefix whaka- can combine with bases such as pau or pakaru: ka whaka.pau.tia ‘(is) used up’. Pau denotes the state of something that was used up, while whaka.pau.tia means the action of something being used up.
b. Nominalized
te patunga o te mango e te tamaiti
\textit{DET kill.NMLZ of DET shark by DET child}
\textit{‘the killing of the shark by the child’}

In example (5.66b), although the passive -\textit{Cia} is absent, the passive interpretation is implied by the use of \textit{e ‘by’}, and the genitive marker \textit{o} used for marking the patient subject in the passive construction.\textsuperscript{45} -\textit{Canga} nominalizations of passive verbs are accepted by all competent speakers, but the nominalizing suffix is supposed to be attached to the stem, as in (5.66b), not to the passive suffix. However, double suffixes are found in the data compiled by Boyce (2006),\textsuperscript{46} such as the example in (5.67), albeit the occurrence is rare.

\begin{align*}
(5.67) \quad & I \text{ mua } \text{ o } te \text{ whaka.tuw.heratangahia o te marae nei } \text{nā…} \\
& \text{P before of DET cause.open.NMLZ.PASS of DET marae here there} \\
& \text{‘Before the opening of these marae…’}
\end{align*}

\textit{Whakatuwheratangahia} contains three affixes: the causative prefix \textit{whaka}-, the nominalizing

\begin{flushright}
\textsuperscript{45} Although the presence of the agent marker \textit{e} and the use of the genitive \textit{o} enable the passive interpretation, they do not necessarily serve as proof that the passive suffix -\textit{Cia} was present before -\textit{Canga} was attached. The agent marker \textit{e} can be used without -\textit{Cia} as in \textit{I pau katao ngā rare te kai e ngā tamariki} (TAM used up all DET lollies DET eat by DET children) ‘The children ate all the lollies.’ (Harlow 2001:177), and the \textit{o} can also be used to mark the non-agentive subject in the nominalized phrase as in \textit{nō te kite.nga ake o te tangata whenua i te ope} (P DET see.nmlz up of DET tangata whenua DO DET party) ‘when the tangata whenua saw the visiting party’ (Harlow 2001:209).
\textsuperscript{46} These examples are from the data collected by the Mary Boyce Corpus (MBC). MBC consists of over one million words, which are taken from recordings of Māori radio and television broadcasts, collected in 1995 and 1996. There are 273 texts of varying lengths and sources. Each text is identified with the date of the broadcast, the program, the topic, the speakers, and the speakers’ tribal affiliation.

In the MBC, I have identified thirty words that have two or more suffixes. Of the thirty words, there are twenty-one that show nominalization followed by passivization. Six cases show passive suffixes with different forms used with nominalization. The remaining three words have the nominalization following the passivization. Of the thirty cases, twenty-two were uttered by the same speaker, who is from Ngāi Tūhoe, in the eastern part of the North Island. Due to limited contact with Western settlers, Tūhoe people are known for their strong Māori identity and strong language background. The speaker of this particular data is a native Māori speaker. Per personal communication with Boyce, it became clear that this double suffixation is a well-accepted usage among Tūhoe speakers. Indeed, for the remaining eight occurrences, only three speakers are not from Ngāi Tūhoe, and one of the three is from Ngāti Awa, a neighboring area to Tūhoe. It is highly likely that speakers in Ngāti Awa share similar characteristics with Tūhoe speakers. One of the remaining two speakers is from Taranaki, the opposite side of the North Island from Tūhoe, while the other is from Tai Tokerau, a stretch of land from Auckland north to Cape Reinga at the very top of the North Island. Therefore, it is quite safe to say that this double or triple suffixation is a productive word formation method in certain dialects.
\end{flushright}
suffix -tanga, and the passive suffix -hia, all of which are attached to the verb tuwhera ‘open’.

As in (5.66b), the use of o in the nominalized phrase suggests that the marae in (5.67) is the patient subject in the passive construction. This suggests that -hia in (5.67) is used for the same function as in (5.66), although the suffix -hia is one of the default forms, as it attaches to the nominalizing suffix tanga and thus no thematic consonant from the stem is accessible.

There are cases of -Canga attached to -Cia:

(5.68) …i mua o te kati.a.tanga o ngā hōhipera puta noa
      P before of DET shut.PASS.NMLZ of DET hospital throughout
      ‘…before the hospital’s being shut.’

There are also cases of -Cia used twice, before and after -Canga:

(5.69) te patu.a.tanga.hia o te tumuaki o tēnei
       DET kill.PASS.NMLZ.PASS of DET headmaster of that
       kura e tētahi tamaiti…
       school by a child
       ‘the headmaster of that school being killed by a child…’

Note that the two passive suffixes take different forms: The first passive suffix is -a, which occurs immediately after the stem patu, and the second passive suffix is the default form -hia, which follows the nominalizing suffix -tanga. Note that the first passive suffix -a in patu.a.tanga.hia is a lexically conditioned allomorph due to the final consonant that was present in Proto Oceanic, but dropped in Proto Polynesian, as explained in chapter 3 (footnote 10), whereas both -tanga and -hia are the default forms. As mentioned earlier, the default forms are used when no thematic consonant is available at the end of the stem. The phenomenon of double suffixation provides supporting evidence that the so-called “passive” -Cia is used more for semantic reasons (e.g., affectedness, perfectiveness, marking high transitivity) than for marking the passive construction. It is plausible to nominalize a passive sentence, as in (5.68), but
passivization of a nominalized item is conceptually unlikely. In particular, -hia in example (5.67) must be in use for a type of emphasis, that is, a semantic reason,\(^{47}\) as one suffix should be enough for the derivation of a type. This, in turn, suggests that double suffixation is best considered a process distinct from the derivational process just described. I suggest that, unlike the types -CANGA and -CIA that attach to types such as PATU ‘HIT’ and INU ‘DRINK’, for instance, -Canga and -Cia in double suffixation attach to the tokens. If it is a derivational process to create a new lexical item, the suffixes need to be accessible to the thematic consonant to derive the correct form. As described in the discussion of example (5.66), the two suffixes need not co-occur, as the passive meaning can be expressed by the use of e to mark the agent and the use of o as the choice of possessive particle in the nominalized phrase. Thus, the use of more than one suffix is an “extra” element that appears to be present for semantic or prosodic reasons.

The double suffixation phenomenon is rare, and more investigation is necessary. However, it does provide evidence that a semantic account of -Canga and -Cia offers a more uniform analysis than the analysis of -Canga as a nominalizing suffix and -Cia as a passive suffix.

5.11. Conclusion

In this chapter, I have presented a possible analysis of Māori as a precategorial language. Māori is similar to Tongan in that its lexical items can appear with either TAM or DET with relative flexibility. I have shown that if we suppose that the bases in phrasal nuclei are types before they are inserted in the nucleus slot in the phrase, then all Māori bases are potentially

---

\(^{47}\) It is possible that -hia is added for some kind of prosodic reason, as suggested by Boyce (personal communication).
compatible with either TAM or DET, and if there is incompatibility between a preposed particle and a base, it is not because the base belongs to a different class than those that are compatible. The types exist, carrying only the concept that the word contains, and the meaning of the word as a token is determined by the meaning of the corresponding type in conjunction with the meaning of the particle to some extent, and by contexts. Following this reasoning, Croft’s prototype analysis cannot tell much about Māori words, because the Māori item without the phrasal category (DET or TAM) is just a concept about the word, not an actual word that is used in syntax.

As Peterson (2013) points out, Māori grammar can be explained in terms of the noun/verb/adjective distinction. However, there is no need to explain Māori word classes in these terms, and an explanation that does not refer to this distinction offers a much simpler account. As the one issue that is agreed on among the scholars of Māori grammar is that the basic grammatical unit in Māori is a phrase, it makes a great deal of sense to search for an account that supports this premise. Māori is a phrase-based language, which has two phrasal categories, and the grammar does not require lexical categories: this is the main fact for Māori word classes.
6. Conclusion

This investigation into Māori grammar started with the simple observation that Māori has the word category overlap phenomenon. I found it interesting that in my first and second year Māori language classes, this phenomenon was not a big issue among my fellow students, and the instructor never addressed it as something that needed special attention or as a unique Māori feature. We simply learned that kei te and i te are for verbs, while te/ngā are for nouns. Even when we found te mōhio ‘the know’ in a reading passage, nobody questioned how the determiner and verb could co-occur. This absence of discussion might be because English speakers are used to zero-derived words, such as cut (v): cut (n) and mail (v): mail (n), and they assumed that in Māori, mōhio is that type of word. Or it may be because the students were oblivious to such details, as they were not used to paying attention to grammatical issues. It may be because we are braced to encounter different and unique grammatical rules when we learn a foreign language. Or it may be because it did not matter whether mōhio is a verb or a noun, because all that mattered to us was to learn how to say what we wanted to say correctly. The category overlap issue was never an issue in this language class, except for one incident that I can recall. A student did not know how to translate an English sentence that contained an adverb, “Paddle strongly,” into Māori. At that time we had not yet learned how to translate such sentences, but I suggested Kia kaha te hoe (TAM strong DET paddle). Genuinely impressed, the other student asked me how I came up with the correct answer, which she said she could have never thought of on her
own. “I couldn’t find strongly in my Māori dictionary. Don’t they have adverbs in Māori?” she asked. At this moment, the category had become an issue.

It is often taken for granted that every word is assigned to a part of speech. Even elementary school students are taught this. *The students do not like difficulty homework* is ungrammatical because difficulty is not an adjective in form and cannot modify a noun in the prenominal position. But *The students do not like math homework* is grammatical even though math is not an adjective, because math homework is one word (a compound noun). *The students did good* should be *The students did well*, because good is an adjective, which is a modifier of nouns, and cannot modify a verb; the correct adverb is well. Our school grammar lessons revolve around such part-of-speech distinctions, which extend into the study of linguistics. Words are assigned to parts of speech and they can occur in the slot specified for their part of speech, and that is how correct sentences are formed. While this is stated here in rather plain language, it is basically the stance that generative grammar takes. The lexical categories are inherently given to the lexical items, and they are syntactic primitives, which operate on the syntactic rules that generate the proper configurations between the constituents: In the beginning was the Word, and the word was with a category.

Therefore, upon hearing that “the categories overlap in Māori,” one naturally asks the question, “Does Māori have parts of speech?” In order to answer this question, we must define the parts of speech, and then the next question is whether the parts of speech are universal. As the investigation continued, I found more questions than answers. Meanwhile, I started to wonder if we were asking the right questions. When I revisited the original issue, that “Māori word categories overlap,” I questioned this statement itself: What kind of categories are they? That is, why are we using categories that overlap? Biologist Stephen Jay Gould writes:
Taxonomy (the science of classification) is often undervalued as a glorified form of filing—with each species in its folder, like a stamp in its prescribed place in an album; but taxonomy is a fundamental and dynamic science, dedicated to exploring the causes of relationships and similarities among organisms. Classifications are theories about the basis of natural order, not dull catalogues compiled only to avoid chaos. (Gould 1989:98)

Searching for nouns, verbs, and adjectives in a language does not fulfill the same goal as searching for categories for a grammatical description of a language under analysis. If we apply Gould’s view of classification, the categories in Māori should be dedicated to exploring the causes of relationships and similarities among the words. Therefore, if the categories overlap, it implies that either the language can allow the categories to overlap because it has a very special grammar that accommodates the overlapping categories (e.g., a very fixed word order requirement), or that we are using the wrong categories. It is possible that languages like Māori that have “overlapping categories” might undermine our confidence in the parts of speech and the grammatical theories that depend upon them, because the questions these languages provoke—“What kind of categories are they?” and “Why are we using categories that overlap?”—take us in a different direction: “What are the categories for?” and “What kind of categories should they be for Māori?” Gil (2000:191) makes this point precisely: “What are the significant syntactic patterns in the language, and what are the categories that must be posited in order to enable the necessary generalizations to be stated?”
I showed in chapter 2 that the concept of parts of speech originated with Classical Greek, a language in which the words could be divided into groups thanks to the elaborate inflectional systems that served to identify nouns and verbs. Therefore, the parts of speech were originally a morphology-centered, language-specific word classification system. The system was inherited by Latin grammarians and other European scholars, enduring for centuries. It was the American structuralists who noticed that the part-of-speech system was not applicable to the non-European languages they were attempting to describe. They made a conscious effort not to resort to their preconceived ideas, including word classification, when they faced unknown languages. Because the languages they dealt with were not always equipped with inflectional and derivational morphemes to mark the categories, they had to rely on the distribution of the constituents. Syntax thus became the important criterion. When generative grammar became the mainstream theory in the field of linguistics, syntax was at the center of the theory, and the parts of speech were its core. The syntactic rules are written in mathematical-looking notations, such as “S → NP □ Aux □ VP”, in which the category notations N, V, and so on, are the crucial part of the syntactic rules. “N” and “V” are not the same as the nouns and verbs of the classic part-of-speech system; they are syntactic categories that inherently belong to the lexical items in the lexicon, and that exist in all human languages. Although the categories are crucial in generative theory, it is not interested in how lexical items are stored in the lexicon for syntax, nor with how the categories are defined. Some linguists within the generative grammar tradition, however, have been interested in defining the lexical categories (which they consider universal), category assignment strategies, and the role of semantics in the generative framework. Meanwhile, cognitive grammar, which views language as one of the cognitive activities we human beings perform, proposed the alternative view of the prototype categories, instead of the classic categories. Croft (1990, 2001,
2004) constructed a table to map out the relationship between three propositional acts (reference, modification, predication) and three semantic classes (objects, properties, actions). The table explains what the things that we have been calling “nouns,” “verbs,” and “adjectives” are and why we are inclined to conclude that all languages should have at least nouns and verbs: they are the prototypes (that are universal). Universality in languages is an intriguing question. If there is anything universal among the human languages, shouldn’t the parts of speech, which have over two thousand years of history in language studies, be a good candidate?

Indeed, categories should exist. If two words show different distributions or attach to different affixes, these two words belong to two different categories. Although Gil (2000, 2013) introduced to the world a language with only one open category (and claimed that others like it exist), most languages have categories. Categories should exist as a part of a language because they are a contributing part of the grammatical system. To the question of whether Māori has parts of speech, I would answer yes, if what is meant by parts of speech is Croft’s prototype categories. Māori has prototypical nouns, verbs, and adjectives, like any other language. But these prototypes are not contributing members of Māori grammar, because the nouns, verbs, and adjectives overlap, which means they are not functioning as they should if they were a part of the grammar. In considering this issue, I have introduced several approaches various linguists have taken to account for the category overlap phenomenon in other languages. Linguistic investigation has revealed that some languages show the nominal and verbal distinction at the syntactic level only. Some languages may have a completely different classification system, which is only relevant on the syntactic level. As mentioned earlier, there is only one known example (although others may exist) of a language that has only one category. Nevertheless, a single example is sufficient to make us consider the existence of such languages. Many
researchers have taken a stance that is diametrically opposed to the general belief and the principles of generative grammar: Lexical categories are not relevant to some languages, lexical categories are not an inherent part of lexical items, and the nominal–verbal distinction does not exist in all languages.

I have entertained the possibility that Māori is one such language that does not have preclassified lexical items or a nominal versus verbal distinction. Māori linguists all concur that the basic unit of grammar in Māori is not a word, but a phrase, and that a word with lexical meaning needs to be part of the tripartite structure of a phrase to be used in a sentence, which means that a Māori word needs at least a phrase marker and possibly a postpositional particle to be a fully functioning word. These facts imply that Māori words are category-free until they are inserted in a phrase that has a syntactic category. Inspired by Broschart’s account of Tongan, I have tried to give a similar account of Māori, which would provide a simpler explanation of Māori word classification. I have come to the conclusion that Māori can be considered one of the type/token languages, and that Māori lexical items (bases) are category-free: They are types, which exist in the lexicon as concepts, not as words to be used in sentences. Only by being part of TAM and DET phrases can they function as words. On the other hand, although TAM is usually strongly associated with verb/predicate and determiners are usually associated with nominal/reference, and they form two separate paradigms in nominal/verbal languages, Māori TAM and DET are not two different entities with separate domains. Rather, they are on a continuum. They share the same slots instead of each forming its own list of the base categories that can follow it, and they share the same distributional tendencies from the syntactic point of view. Thus, Māori has two phrase categories; the bases are flexible in attaching to either TAM or DET, and incompatibility is due to semantic constraints.
As Peterson (2013) emphasizes, both accounts, with noun/verb/adjective distinction or without n/v/a distinction, are possible to account for the language of his interest (Kharia), and he concludes that the simpler account should be chosen on the principle of Occam’s razor. I would further suggest that the choice is also a matter of purpose. Are we analyzing Māori to make a grammatical description of it, or are we using Māori as an example, or a tool, to prove a particular theory? When both accounts work, researchers with the first purpose will choose the simpler account, while those with the second purpose will choose the one that supports the theory. Especially in regard to the topic of this dissertation, upon seeing how generative grammar that Māori has the noun/verb/adjective distinction, my reaction is that it works because the theory assumes from the beginning the existence of N, V, and A. It is thus made to work, albeit with many stipulations.

During the course of this investigation, some interesting facts emerged. First of all, I have observed the strong influence of English on the analysis of Māori. American structuralists warned about this unfortunately inevitable aspect of language investigation. It is impossible to rid ourselves of all the knowledge we have acquired from our first and second languages when we analyze a language new to us. However, it is possible to be constantly aware of this pitfall so that the hindrance or contamination will be minimal. Before giving a gloss of TAM for he in the predicate clause and ki te in the infinitival clause, the glosser should ask if such an interpretation is possibly motivated by the English translation. That is what the American structuralists campaigned for: that languages be analyzed within their own systems, by their own grammatical structures, in their own terms.

Second, this investigation has drawn out the interesting nature of the denominal verbs. In Māori, the so-called verbs all appear with DET, but not all so-called nouns can appear with
TAM. This is the reason why we are tempted to set up a special class for nouns. Even Biggs did so. However, the fact that most nouns remain nouns while verbs become nouns freely is not because there are two types of nouns, one that can appear with TAM and one that cannot. The one that can appear with TAM comprises nouns that developed from verbs in response to a need in the past had the need in the past for the denominal verbs. I believe that this tendency is cross-linguistic. For instance, recently introduced words in English or Japanese include many nouns, which denote new products or ideas, such as e-mail, selfie, and tweet; two of these items have verbal usages. Japanese created the verb sutaba-ru ‘go to Starbucks or meet at Starbucks over a cup of coffee’ from a noun sutaba ‘Starbucks’. Yet in a list of newly created vocabulary, there would be very few newly created verbs. Those that exist are most likely to be denominal verbs. Since we came up with a new mailing method called “e-mail” (n), we can “e-mail” (v). What began as Send me an e-mail quickly became E-mail me. Lieber (2004:27) writes:

In some sense SUBSTANCES/THINGS/ESSENCES have ontological priority over SITUATIONS…SITUATIONS presuppose participants or arguments, which are usually SUBSTANCES/THINGS/ESSENCES do not presuppose situations. In some sense that semantic category SUBSTANCES/THINGS/ESSENCES is prior to, more fundamental than, and necessary for the semantic category SITUATION.

In this sense, I have come to realize that SUBSTANCES/THINGS/ESSENCES, which correspond to so-called nouns, hold a significant place in language. There are probably far more “things” or “objects” in our world that we can or need to derive verbs from. However you name them, for “nouns” or “referents” or SUBSTANCES/THINGS/ESSENCES to be used in verbal
usages, there must be a specific motivation for the denominalization, whereas when so-called verbs are nominalized, it is simply because the constructions in which they appear require them to be used referentially. Whether they are “nouns” or “references” or “arguments,” all languages might actually have this category at a certain grammatical level. For Māori, such a category probably comprises the DET phrases such as te ika ‘the fish’, te kiore ‘the mouse’, and te kaputī ‘the cup of tea’, which still only select DET for the preposed particle.

Third, the concept of type/token languages developed by Broschart (1997) is truly intriguing. Although at this stage, the concept still refers only to what Broschart proposed, that is, the type of languages that are like Tongan, I am inclined to believe that the type/token difference might exist cross-linguistically, but how type and token are realized in actual lexical items is language specific. For example, compare chair in English, isu in Japanese, and tūru in Māori. They all refer to the same object, a ‘chair’ we sit on. However, English chair is the only one that has a verbal usage (*isu-te kudasai ‘please sit down’, *ka tūru ia ‘he/she sits). Moreover, chair (v) does not mean ‘to sit’. It has subcategorization ability, and can take two arguments: I was chosen to chair the new committee. The other two ‘chairs’, in Māori and Japanese, do not have this function. In light of Clark and Clark’s (1979) account of the nature of denominal verbs and the “referential superiority rule,” it is quite interesting that the same concept, ‘chair’, is realized in actual usage with different functions and meanings. Certain pairs are semantically predictable, such as dance (v): dance (n), but others require specific cultural and situational motivations, which Clark and Clark describe as “innovation.” This is why I believe Evans and Osada’s (2005) strict criteria are unrealistic. At least at an empirical level, category flexibility cannot always be bidirectional, strictly compositional, and completely exhaustive throughout the lexicon.
Although Māori word classification is best accounted for by a category underspecification approach, the parts of speech, that is, the noun/verb/adjective system or the prototype categories, are unlikely to lose their solid status in language studies, and they have their purposes. They do exist in some languages as grammatical categories, not just as cognitive categories. It is also useful to have a common terminology for the grammatical description of diverse languages. For instance, labeling the categories Class A, Class B, and so on in language X and Class 1, Class 2, and so on in language Y would not capture the similarities that might exist in the two languages at a glance. Well-used terms such as “noun” and “verb” have a definite advantage in this respect. However, when one needs to document a hitherto undocumented language, it is best not to make a list of “nouns” and “verbs.” Gil (2000:191) writes, “most grammatical descriptions automatically assign words and phrases to syntactic categories on the basis of their meanings.” Calling a word a “noun” means more than that it is a thing/place/name. It specifies the word’s grammatical function, as a reference used for an argument. Therefore, the “labeling” of words with the terminology of nouns and verbs can block us from seeing the language under investigation in any other way but as a nominal/verbal language.

Understanding the difference between Māori word categories and English word categories should be helpful to Māori instructors, textbook writers, dictionary makers, and learners. Labeling vocabulary items as “nouns” and “verbs” does not help students who are learning Māori. My Māori classmate could not translate a sentence with an adverb into Māori, because the materials she had access to provided no way for a non-Māori speaker to know that she had to say “the paddling be tough” to mean “Paddle hard!” It makes sense to introduce new lexical items to learners in their actual usage, that is, in a phrase. While this can be said for all
second language learning situations, it is the only way one can learn Māori correctly. As a teaching and learning strategy, this is not merely a matter of convenience. It is based on the legitimate claim that Māori is a phrase-based language, in which the phrase is the basic unit of grammar, and a lexical item by itself is only a concept.
Bibliography


171


