Studies in Constraint-Based Lexicalism

A series edited by
Miriam Butt, *University of Konstanz*
Tracy Holloway King, *Xerox Palo Alto Research Center*
Jean-Pierre Koenig, *State University of New York at Buffalo*

The aim of this series is to make work in various nonderivational, lexicalist approaches to grammar available to a wide audience of linguists. In approaches of this kind, grammar is seen as the interaction of constraints from multiple dimensions of linguistic substance, including information about syntactic category, grammatical relations, and semantic and pragmatic interpretation.

Studies in
Constraint-Based Lexicalism

Voice and Grammatical Relations in Austronesian Languages

edited by
Peter K. Austin and Simon Musgrave

CSLI
PUBLICATIONS
Center for the Study of
Language and Information

Center for the Study of Language and Information
Leland Stanford Junior University
Printed in the United States
12 11 10 09 08 1 2 3 4 5

Library of Congress Cataloging-in-Publication Data

Voice and grammatical relations in Austronesian languages /
edited by Peter K. Austin and Simon Musgrave.

p. cm. – (Studies in constraint-based lexicalism)

Includes bibliographical references and index.
ISBN-13: 978-1-57586-500-3 (pbk. : alk. paper)
ISBN-10: 1-57586-500-9 (pbk. : alk. paper)
ISBN-13: 978-1-57586-499-0 (cloth : alk. paper)
ISBN-10: 1-57586-499-1 (cloth : alk. paper)
1. Austronesian languages—Voice.
2. Austronesian languages—Verb.
3. Austronesian languages—Syntax.
I. Austin, Peter K. II. Musgrave, Simon, 1959–

PL5034.V65 2008
499'.2–dc22 2007051284
CIP

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

CSLI was founded in 1983 by researchers from Stanford University, SRI International, and Xerox PARC to further the research and development of integrated theories of language, information, and computation. CSLI headquarters and CSLI Publications are located on the campus of Stanford University.

CSLI Publications reports new developments in the study of language, information, and computation. Please visit our web site at
http://cslipublications.stanford.edu/
for comments on this and other titles, as well as for changes and corrections by the authors, editors, and publisher.

Contents

territory that belongs to the applicative, whereas the reverse is not true.[6]

From these observations we must conclude that the causative must be the least marked of the valency-increasing devices.

7 References

Alsina, Alex. 1992. On the argument structure of causatives. *Linguistic Inquiry* 23:517–555.

Alsina, Alex. 1993. Predicate composition: a theory of syntactic function alternations. Ph.D thesis, Stanford University.

Alsina, Alex. 1996. *The role of argument structure in grammar: evidence from Romance*. Stanford: CSLI.

Austin, Peter. 1997. Causatives and applicatives in Australian Aboriginal Languages. In *Dative and related phenomena* ed. Kazuto Matsumura & Tooru Hayasi. 165–225. Tokyo: Hitsuji Shobo.

Baker, Mark C. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.

Donohue, Mark. 1996. Relative clauses in Tukang Besi: grammatical functions and thematic roles. *Linguistic Analysis* 26.3–4:159–173.

Donohue, Mark. 1997. The applicative construction in Tukang Besi. In *Proceedings of the Seventh International Conference on Austronesian Linguistics* ed. Cecilia Odé & Wim Stokhof. 415–432. Amsterdam: Editions Rodopi B.V.

Donohue, Mark. 1999. *A Grammar of Tukang Besi*. Grammar Library series No. 20. Berlin: Mouton de Gruyter.

Harris, Alice & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge, UK: Cambridge University Press.

Lefebvre, Claire. 1991. *Serial verbs: grammatical, comparative, and cognitive approaches*. Amsterdam: John Benjamins.

Rugemalira, Joseph M. 1994. The case against the Thematic Hierarchy. *Linguistic Analysis* 24:62–81.

[6] That is, there are no cases of causative morphology being associated with applicative functions, something like:

(iii)	*No-pa-[verb]='e*	*na*	*iai=su*
	3RL.SUBJ-CAUS-[verb]=3.P	NOM	younger.sibling=1SG.GEN
	'They [verb]ed for my younger sister.'		

The only reading possible for a sentence of the type seen in (iii) would be 'They made my sister [verb].'

10

Lexical Categories and Voice in Tagalog

NIKOLAUS P. HIMMELMANN

1 Abstract

The meaning and lexical category of so-called verbal roots is one aspect of the voice systems in Tagalog and other Philippine-type languages which has received little attention in the controversy surrounding the analysis of these systems. It is common to assume that these roots cannot occur without any affixation and that, therefore, they should be considered precategorial. Here it is shown that this view is ill-conceived. To begin with, it is possible to distinguish different classes of roots based on morphological features. Therefore, roots are not precategorial. Furthermore, a large majority of the putative verbal roots allows for unaffixed uses. However, 'verbal roots' have 'non-verbal' meanings when used without voice marking. Inasmuch as it can be shown that voice-marked forms have clearly 'verbal' meanings, it follows that voice marking is derivational (among many other things).

Voice and Grammatical Relations in Austronesian Languages.
Simon Musgrave and Peter Austin

The first part of the paper is taken up by the discussion of some basic issues regarding the nature of lexical and syntactic categories, which is a prerequisite to sorting out the Tagalog facts. Contrary to standard assumptions, it is proposed that a clear-cut distinction should be made between lexical and syntactic (i.e. phrase-structural) categories, allowing for mismatches between these two kinds of categorisations of lexical items.[1]

2 Introduction

Looking through a standard Tagalog-English dictionary such as Panganiban (1972), Santos (1983), English (1986), or Rubino (1998a), one will notice almost immediately that hardly any Tagalog root is glossed with an English verb. Instead, almost all presumably verbal roots are glossed with English nouns or adjectives/participles. Typical examples include object nouns such as 'gift' for *bigáy*, action nominalisations such as '(act of) crying' for *iyák*, and adjectives/participles such as 'surpassed, defeated' for *daíg*. This practice of glossing Tagalog roots with English nouns or adjectives is in marked contrast with the practice of glossing voice-marked formations involving the same roots as English verbs. For example, *i-bigáy* is glossed as 'to give something to someone, to hand in', *um-iyák* as 'to cry', and *daig-ín* as 'to outdo, to surpass'. Taken at face value, these differences in the treatment of roots and voice-marked formations appear to imply that in Tagalog all verbal expressions are somehow derived from non-verbal roots.

I argue in this paper that there is some truth to the idea that all Tagalog verbal expressions are derived to some degree. A proper and testable explication of this idea, however, requires certain preliminary clarifications of basic structural aspects of Tagalog morphosyntax and the nature of lexical and syntactic categories. Section 3.1 provides a brief overview of the most important structural positions in a Tagalog clause and the terms used here to refer to these positions.

In section 3.2 it is proposed that there is no necessary correlation between the classification of lexical items based on morpho-syntactic features and the classification of lexical items with regard to the slots they may occupy in a phrase structure tree. That is, there is a difference between terminal syntactic categories (the categories of the terminal nodes in a phrase structure tree) and lexical categories proper (i.e. the formal categories to which lexical items may belong, with the exception of the categories of phrase structure), and mismatches between these two kinds of categorisations of lexical items may occur. This proposal is consistent with work challenging the categorial uniformity hypothesis (Bresnan 1994, Bresnan 2001) which argues that what is usually assumed to be just one type of structure (called either lexical or syntactic category) should in fact be dealt with on at least two different levels, and that lexical insertion is not constrained by a one-to-one correspondence between the two. My proposal is supported by detailed examination of lexical facts, in contrast to the evidence put forward by Bresnan which relies on similarities of distributional behaviour of items belonging to different lexical categories.

[1] Many thanks to Gary Palmer, Malcolm Ross, Carl Rubino, Hans-Jürgen Sasse, Eva Schultze-Berndt, Angela Terrill, and John Wolff for very useful comments on a draft version of this paper.

Section 3.2 also introduces a number of terminological distinctions crucial to a proper handling of the morphosyntactic and semantic differences found among lexical items. In particular, an attempt is made to avoid the use of the multiply ambiguous terms *noun* and *verb* since the three levels to which these terms may apply (the ontological, the morpho-lexical, and the syntactic (phrase-structural) level) are not at all commensurate in a language like Tagalog.

Section 4 applies the distinction between (terminal) syntactic categories and lexical categories proper to Tagalog. With regard to syntactic categories it may be argued that all Tagalog content words (both roots and derived words) are categorially indistinct, i.e. they may all occur in essentially the same basic syntactic positions (section 4.1). With regard to lexical categories, however, there are clear-cut categorial distinctions (section 4.2). That is, Tagalog roots exhibit different formal (in particular morphological) properties which are not directly predictable from their meaning. Hence, it is highly questionable whether Tagalog roots can be characterised as precategorial.

Section 5 then turns to the issue of what kind of meanings are denoted by Tagalog roots. It is shown that all kinds of roots can be used without further affixation and that they have a consistent and clearly identifiable meaning in their unaffixed uses. In particular, it is not correct to claim that so-called verbal roots generally do not occur without further affixation.

Finally, section 6 provides an explication of the view that voice marking in Tagalog is derivational in all its manifestations, based on the fact that it changes the meaning and the category of the roots to which it is attached. Furthermore, although the class of voice-marked words in Tagalog shares some similarities with verbs in English, it would be wrong to attribute the same kind of essential properties to both classes of lexical items.

3 Preliminaries

3.1 Basic Syntactic Functions in Tagalog

This section briefly introduces the following four basic morphosyntactic functions in Tagalog clauses: predicate, subject, non-subject argument or adjunct, and modifier. The first three functions are clearly defined in Tagalog by a set of grammatical markers and word order. The definition of the modifier function in addition requires reference to the semantics of the items involved.

The following discussion is confined to the most basic and simple construction types and should not be mistaken for a comprehensive outline of basic Tagalog clause structure. In particular, the fact that syntactic functions are defined primarily in terms of overt grammatical markers and word order should not be misconstrued as the claim that these surface structural phenomena are the only evidence for these functions. Instead, a substantial number of further syntactic facts such as relative clause formation, topicalisation, control phenomena, etc. could be adduced to support the analysis of these functions. The purpose of the present section, however, is simply to provide the reader with a set of easily identifiable features for each function and thus to facilitate the parsing of the examples throughout this paper.

As is well-known, considerable controversy surrounds the question of whether the grammatical relation *subject* exists in Tagalog. Following Schachter (1976), it is generally agreed that *ang*-phrases in post-predicate position (see below) show many but not all of the presumably universal subject properties proposed by Keenan (1976). Still, as argued in detail by DeWolf (1979:67-86, 1988:144-150) and Kroeger (1993), *ang*-phrases may be analysed as subjects because they exhibit a substantial number of important subject properties (such as being the only argument that can launch floating quantifiers, control secondary predicates, be relativised and be omitted in conjunction reduction) while other subject diagnostics are inapplicable or inconclusive. The major point of contention pertains to the so-called agent-related properties of subjects, in particular the properties of serving as the antecedent in reflexive constructions, the target in Equi-NP deletions and the addressee in imperatives. To some extent, Kroeger and Schachter disagree here about the empirical facts (cf. Kroeger 1993:36-40, 71-107 and Schachter 1996:21-27). More importantly, it is doubtful whether these properties in fact provide reliable diagnostics for grammatical relations. Artawa and Blake (1997:505f), among others, profess serious doubts in this regard and argue for the viability of the subject notion in Balinese, a language for which the basic facts relevant to this issue are quite similar to the Tagalog ones (see also Arka (this volume), Arka and Manning (this volume)). Here I adopt the position that there are subjects in Tagalog, with the proviso that the subject in Tagalog differs in some regards from subjects in other languages such as English.

Predicates in Tagalog typically occur in clause-initial position, as in the following example:

(1)	***dumatíng***	*yung*	*asawa*	*niya*
	um-datíng	iyón.LNK	asawa	niyá
	AV-arrival	DIST.LNK	spouse	3SG.POSS

'Her husband arrived.'

If another constituent precedes the predicate, the marker *ay* occurs in front of the predicate:

(2)	*silá mag-iná*	***ay*** *natulog*	*na*
	silá mag-iná	ay na-tulog	na
	3PL RCP-mother	PM RL.STAT-sleep	now

'The mother and her daughter fell asleep.'

Subjects generally follow the predicate. They are marked with the specific article *ang* (in the case of proper names *si*):

(3)	*masaráp*	***ang pagkain***
	ma-saráp	ang pag-kain
	STAT-satisfaction	SPEC GER-eating

'The food was good.'

Note that the specific article *ang* can be replaced by a demonstrative pronoun (such as *yung* in *(1)* above) or, much more rarely, a personal pronoun (instead of *silá mag-iná* in *(2)* one could also say *ang mag-iná*).

Furthermore, the specific article *ang* clearly is not a subject marker since *ang*-phrases may also occur in other syntactic functions. In the following example, the first *ang*-phrase (*ang langgám*) functions as the predicate while the second *ang*-phrase (*ang tumulong sa mga bata'*) is the subject:

(4) ***ang*** ***langgám*** *rin* *ang* *tumulong* *sa*
ang langgám din ang um-tulong sa
SPEC ant also SPEC AV-help LOC
mga *bata'*
mangá bata'
PL child
'The ants also helped the children.'
(lit. The ones who helped the children were also the ants.)

Therefore, it is not possible to define the subject simply as the phrase marked by *ang*. Instead, the subject is defined as the *ang*-phrase which follows the predicate (and there can be only one *ang*-phrase after the predicate).

Although a subject is always implied in a Tagalog basic clause, this subject does not have to be overtly expressed. Compare the following sequence of two clauses where the subject of the second predicate (*inilagáy*), which is coreferential with the subject of the first predicate (*kinuha*), is not overtly expressed:

(5) *at* *kinuha* *niyá* *ang* *langgám*
at in-kuha niyá ang langgám
and RL(UNDR) -getting 3SG.POSS SPEC ant
at ***inilagáy*** *niyá* *sa* *pampáng*
at in-i-lagáy niyá sa pampáng
and RL(UNDR)-CV-position 3SG.POSS LOC river.bank
'And he got the ant and put it on the riverbank.'

If the predicate is voice-marked (as in *(1) and (5)* above), a special relation exists between the subject and the predicate in that the semantic role of the participant appearing in subject function is overtly marked by the voice affix on the predicate. If the predicate is marked with *-um-*, *mag-/nag-*, or *maN-/naN-*, the subject bears the actor role, as shown by *(1)* and the following examples:

(6) ***sumakáy*** ***silá*** *sa* *bangká'*
um-sakáy silá sa bangká
AV-passenger 3PL LOC boat
'They got on the boat.'[2]

[2] In this example, a pronoun (*silá*) is used in subject function. Pronouns do not co-occur with the markers which occur with noun-phrases. Thus they are never marked with *ang*. Furthermore, they are second position clitics, appearing immediately after the first constituent of the predicate.

(7) ***sumigáw*** ***yung*** ***anák***
um-sigáw iyón.LNK anák
AV-shout DIST.LNK child
'That child shouted: ...'

(8) ***nag****machinegun* *na* ***ang*** ***eroplano***
nag-machinegun na ang eroplano
RL.AV-machine.gun now SPEC airplane
'The plane machinegunned.'

In some instances the choice of one or the other of the three actor voice affixes conveys semantic differences pertaining to reflexivity, the intensity of the action and the like.[3] In many instances, however, the choice of the actor voice affix is determined by the root or stem (see section 4.2).

Undergoer voice is marked by one of the following three affixes: the prefix *i-*, the suffix *-an*, and the suffix *-in*. Unlike the three actor voice affixes, the three undergoer voices affixes consistently differ with regard to the semantics of the undergoer. Hence it is customary to distinguish at least three undergoer voices in Tagalog. Ignoring several details and complications, it generally holds true that if the predicate is marked with the CONVEYANCE VOICE prefix *i-*, then the subject expresses an argument bearing the semantic role of a displaced theme. Compare:

(9) ***ibinalík*** ***nilá*** ***ang*** ***bata'***
i-in-balík nilá ang bata'
CV-RL(UNDR)-return 3PL.POSS SPEC child
'They returned the child.'

Here the subject (*ang bata'*) is the displaced theme (i.e. the entity viewed as moving) of the event expressed by the predicate (*ibinalík*). The actor is expressed by a possessive pronoun (*nilá*). In addition to the prefix *i-*, the predicate is marked for realis mood by the infix *-in-*, which only occurs in the undergoer voices.

Instruments are also viewed as moving entities and hence marked with the conveyance voice prefix:

[3] See Pittman (1966), Schachter and Otanes (1972:292f passim), and Wolff et al. (1991:113,821f) for exemplification and discussion.

(10) *ipangpùputol* *ko* *na* *lang*
i-pang-RED1-putol ko na lamang
CV-INSTR-RED1-cut 1SG.POSS now only
itóng ***kutsilyo.***
itó-ng kutsilyo
PROX-LNK knife
'I will just cut it with this knife.' (Wolff et al. 1991:367)

The suffix *-an* marks locative voice. In locative voice, the subject expresses a locative argument, understood in a very broad sense. This may be the location at which something happened:

(11) *tinirhán* *ko* ***ang*** ***bahay***
in-tirá-an ko ang bahay
RL(UNDR)-dwelling-LV 1SG.POSS SPEC house
na ***itó***
na itó
LNK PROX
'I stayed at this house.'

Or the location to which (or from which) motion occurred:

(12) *pinuntahán* *na* *namán* *nilá*
in-puntá-an na namán nilá
RL(UNDR)-direction-LV now also 3PL.POSS
ang ***bata'***
ang bata'
SPEC child
'They went to the child.'

Locative voice is also used for recipients, addressees, and benefactees *(13)*:

(13) *tìtirán* *ninyó* ***akó***
RED1-tirá-an ninyó akó
RED1-leftover-LV 2PL.POSS 1SG
'Will you (please) set some aside for me.'

Even more generally, locative voice may be used for all kinds of undergoers which are not directly affected by the action denoted by the predicate:

(14) *hindî'!* *tingnán* *mo* ***si*** ***Maria***
hindî' tingín-an mo si Maria
NEG look-LV 2SG.POSS PN Maria
'Don't (panic)! Just look at Maria!'

(15) *tulungan* *ninyó* ***akó***
tulong-an ninyó akó
help-LV 2PL.POSS 1SG
'(If) you help me, ...'

In *(13)* and (15) the subject is the first person singular pronoun *akó*, in *(14)* it is a proper noun (Maria) which is marked with the proper noun article *si* rather than with the specific article *ang*. Examples *(13)*-(15) also illustrate undergoer voice predicates in non-realis mood, which lack the realis marking infix *-in-*.

The suffix *-in* marks patient voice. It is the unmarked member of the undergoer voice-marking affixes and is used for a broad variety of undergoers, including prototypical patients, i.e. entities directly affected or effected by the event denoted by the predicate:

(16) a. *patayín* *natin* ***itóng*** ***dalawang***
patáy-in natin itó-ng dalawá-ng
dead-PV 1PL.IN.POSS PROX-LNK two-LNK
Hapón
Hapón
Japan
'Let's kill these two Japanese!'

The suffix *-in* differs from the other two undergoer suffixes in that it only occurs in non-realis mood (as in the preceding example). In realis mood, the predicate is simply marked by the realis undergoer voice infix *-in-*:

b. *pinatáy* *natin* *itóng*
in-patáy natin itó-ng
RL(UNDR)-dead 1PL.IN.POSS PROX-LNK
dalawáng *Hapón*
dalawá-ng Hapón
two-LNK Japan
'We killed these two Japanese.'

Recall that the realis infix *-in-* occurs in all, and only, the undergoer voices (cf. examples *(9)*-(12) above).

This brief review of voice marking concludes our introductory discussion of the two basic syntactic functions subject and predicate. Turning now to the third syntactic function mentioned at the beginning of this section, non-subject argument or adjunct, note first that it is quite difficult to make a clear-cut distinction between non-subject arguments and adjuncts in Tagalog. Since the distinction between those two syntactic functions is of no relevance to the issues of primary concern in this paper, it will be ignored here.

Non-subject arguments and adjuncts are marked with either the genitive marker *ng* or the general locative marker *sa* (in the case of proper nouns, the markers are *ni* and *kay*, respectively). Neglecting some minor uses, *ng* primarily marks possessors (as in (17)) and non-subject arguments (as in (18) and (19)):

(17)

ang	*hari*	***ng***	***lamok***
ang	hari	ng	lamok
SPEC	king	GEN	mosquito

'The king of the mosquitos.'

(18)

pùpunuín	*mo*	*iyán*	***ng***	***kuto***
RED1-punó'-in	mo	iyán	ng	kuto
RED1-full-PV	2SG.POSS	MED	GEN	louse

'You fill that (cup) with lice.'

(19)

kinagát	***ng***	***mga***	***langgám***	*ang*	*mama*
in-kagát	ng	mangá	langgám	ang	mama'
RL(UNDR)-bite	GEN	PL	ant	SPEC	man

'The ants bit the man.'

The locative marker *sa* marks a large variety of temporal and local adjuncts *(20)* and recipients/goals, as well as (some) definite patients and themes when they do not occur in subject function (cf. *sa mga bata'* in *(4)* above):[4]

[4] Another common gloss for *sa* is OBL(IQUE). This gloss implies that all *sa*-phrases are syntactically oblique. Though this implication may turn out to be true, it has not yet been explicitly shown that all *sa*-phrases in Tagalog are in fact syntactically oblique (and that they differ in this regard from *ng*-phrases). Therefore, I prefer to use a gloss which leaves this issue open.

(20)

at	*dun*	*na*	*sila*	*tàtabunan*
at	doón	na	silá	RED1-tabon-an
and	DIST.LOC	now	3PL	RED1-complete.cover-LV

sa	***lugár***	***na***	***iyón***
sa	lugár	na	iyón
LOC	place	LNK	DIST

'And there they were covered with earth at that place.'

(21)

nagpunta	*silá*	***sa***	***simbahan***
nag-puntá	silá	sa	simbahan
RL.AV-direction	3PL	LOC	church

'They went to a church.'

A major difference between *ng*-phrases and *sa*-phrases pertains to the fact that the position of *sa*-phrases is flexible (they may occur in pre-predicate as well as in post-predicate positions) while the position of *ng*-phrases is fairly restricted (they generally immediately follow their head).

Turning finally to modifiers, all modifying constructions involve the linker *na* (*-ng* after vowels, /n/ or glottal stop) between the two constituents of a modifying construction:

(22)

an	*malíit*	***na***	*hayop*
ang	ma-líit	na	hayop
SPEC	STAT-smallness	LNK	animal

'the small animal'

The order of the constituents in a modifying construction is not fixed ('the small animal' could also be rendered by *ang hayop na malíit*). Semantically it is in general quite clear which constituent denotes the (semantic) head of the construction and which the modifier. That is, in *(22)* it is clear for speakers of Tagalog that what is denoted is a small specimen of the class of animals and not an animal-like specimen of the class of small things, regardless of the order of *malíit* and *hayop*. However, it is not clear whether this semantic distinction has any kind of formal (prosodic and/or syntactic) correlates.

Consequently, whenever in the following sections it is said that a given word or root can (or cannot) be used as a modifier, the term *modifier* does not refer to a constituent which is defined exclusively in syntactic terms. It is defined syntactically in that the term *modifier* is used only in reference to constituents which are linked to the preceding or following constituent with a linker. Which of the two linked constituents is considered the modifier (and which one the head), however, is decided on semantic grounds.

To summarize: the four basic syntactic functions predicate, subject, non-subject argument or adjunct, and modifier are easily identifiable in Tagalog because there is a set of markers which in combination with a few positional restrictions allows a straightforward identification of each of these functions (with the exception of the modifier function which necessarily involves reference to the semantics of the two items joined by a linker). What is of major importance for the following discussion is the fact that the grammatical markers are distributed in such a way that in principle, each Tagalog content word, except a clause-initial predicate, is preceded by one grammatical marker (or function word). This has important consequences for Tagalog phrase structure, as will be seen shortly.

3.2 Lexical and Syntactic Categories

The nature of syntactic categories (parts of speech) in Tagalog is a matter of controversy. Specifically, there is a long tradition of claims that the distinction between nouns and verbs is minimal or perhaps even non-existent.[5] Many of these claims suffer from confusion in a number of respects, resulting from the widespread practice of not differentiating clearly enough between lexical categories, i.e. the classification of lexical items according to grammatical (phonological, morphological, morphosyntactic) criteria, and syntactic categories, i.e. the category labels attached to the nodes of a phrase structure tree. Although there is an interrelation between these two kinds of categorisation, I follow here the view set out in Sasse (1993a/b)[6] that these two kinds of categorisation should be clearly distinguished and that there is no necessary correlation between them. The present section provides a very condensed version of the argument for this view. It also introduces and exemplifies the crucial distinctions to be used in the analysis of Tagalog in the following sections.

To begin with, a clear distinction should be made between ontological and linguistic (grammatical) categorisation. Ontological (conceptual, notional) categories are the result of the categorisation of the entities populating the universe, as perceived and conceived by the human cognitive apparatus (see, for example, Jackendoff 1983:48 passim). Inasmuch as the human cognitive apparatus is universal such a categorisation is universal (otherwise it is, at least in part, culture-dependent). Lexical and syntactic categories are the result of the categorisation of linguistic items on the basis of grammatical (phonological, morphological, syntactic) criteria. That is, the evidence for lexical and syntactic categories always involves at least one formal property of the items in question (cf. Sasse 1993a:649).[7] Lexical and syntactic categories are thus by definition language-specific as they are based on language-specific formal features and the distribution of such features tends to show language-specific idiosyncrasies (that is, even among closely related languages which share a basic inventory of lexical items and grammatical features, the categorisation of the lexical items based on the grammatical features will not produce classes of lexical items which are fully commensurate).

The view that lexical and syntactic categories cannot be defined in purely notional terms does not imply that there is no interrelation between ontological and syntactic categorisation. Generally, there will be a substantial overlap between the classes resulting from the two different categorisation procedures. A perfect match between an ontological category and a single, semantically well-motivated grammatical feature, however, is suspicious in that it raises the issue of whether the class in question is really a (formal) grammatical category. A typical example is the use of periphrastic comparative and superlative constructions to define the grammatical category *adjective*. If no other features correlate with these constructions, the class established in this way is the class of gradable concepts which, I would hold, is an ontological rather than a linguistic class. In languages where *adjective* is a formally well-defined class it is typically the case that not all adjectives are gradable (e.g. *dead*, *ready*). On the other hand, lexical items which clearly are not adjectives on formal grounds can be used in comparative constructions simply because they are ontologically gradable (e.g. *Where would you find more beauty than in this place?*).

To put this in more general terms, as long as reference to an ontological (notional) category is enough to identify the class of items which partake in a given grammatical construction or show a given grammatical feature, there is no need to employ a second layer of grammatical categories in order to delimit the class of items in question. Such grammatical categories would be simply copies of the ontological ones. Only if there is some mis-

[5] For example, Müller (1882:99ff), Scheerer (1924), and Capell (1964) all claim that Tagalog verbs are not really verbs but nouns. (Capell (1964:244ff) lists a number of further authors making the same kind of claim). Bloomfield (1917) and Lopez (1977) imply through their terminology that there is no grammatical distinction between nouns and verbs. Schachter and Otanes (1972:62) make a distinction between nouns, verbs and adjectives largely for expository purposes. More recently, lack of a distinction between nouns and verbs has been claimed for various levels by Lemaréchal (1982, 1989), Himmelmann (1987, 1991), Gil (1993), and Naylor (1995).

[6] See also Himmelmann (1991:25,44f, 1997:111-124) and Broschart (1997).

[7] See also Anward et al. (1997:172) who make a similar point in a somewhat confusing way. They first assert that 'the primary, definitional properties of parts of speech are semantic or pragmatic, rather than form-related', only to go on to qualify this statement with the constraint that 'semantic or pragmatic features are part-of-speech-defining only if there is at least one formal characteristic that correlates with them'.

match between grammatical and ontological classes is it possible and useful to define grammatical classes of lexical items.

There is no doubt that ontological categories are of fundamental importance to the crosslinguistic study of lexical and syntactic categories. It is only by correlating the formally determined lexical and syntactic categories with ontological categories that it becomes possible to make a crosslinguistic comparison between syntactic and/or lexical categories. That is, in order to claim that two languages, L1 and L2, both have a category *noun* it has to be shown that class A of lexical items in L1 and class A of lexical items in L2, each of which is defined in terms of a set of language-specific features (say, in L1 class A is defined by the inflectional formatives it occurs with while in L2 class A is defined by its co-occurrence with a copula), are similar in two regards: a) they are of roughly equal size; and b) prototypical members of both classes denote ontologically similar entities, including animate beings and (perceptible and time-stable) things.

And precisely for reasons of crosslinguistic comparability, it will occasionally be necessary here to make reference to a number of ontological categories, in particular THINGS, PERSONS, ACTIONS, STATES, and PROPERTIES.[8] The term OBJECT (in small caps) is used to refer to all time-stable entities (in particular ANIMATE BEINGS and THINGS). It is assumed without further discussion that Tagalog ontology is very similar, if not identical, to English ontology so that, for example, a Tagalog root which denotes 'stone' is assumed to be a THING-denoting root.

Turning now to the grammatical categorisation of lexical items, a first and very basic distinction is generally made between content words and function words (or full words and particles).[9] Such a distinction appears to be possible in all languages (cf. Sasse 1993a:652f). Furthermore, it is quite generally agreed that the inventory of function words is highly language-specific. Whenever there is a controversy regarding the number and kind of syntactic categories in a given language, it pertains to the linguistic classification of content words. And with regard to the classification of content words the distinction between lexical and syntactic categories is of central relevance.

[8] The alternative to using ontological categories is to use 'noun', 'verb', 'adjective' in a rather loose ontological sense (i.e. as terms for the ontological categories typically covered by the members of these grammatical classes). This procedure is widespread in the literature but easily leads to the confusion of ontological and grammatical categories and thus is avoided here.

[9] The same distinction is made on a categorial level, i.e. the distinction between lexical and functional categories.

The term *lexical category*, as used in much of the current literature, is applied to two, only partially overlapping, categorisation procedures. In one usage, which is widespread among syntacticians, it refers to the terminal nodes of a phrase structure tree, also known as lexical insertion points. In the transformational tradition it has been widely assumed that the set of major lexical categories universally consists of the following four members: N(oun), V(erb), A(djective), and P(reposition) (or, expressed by way of a feature matrix, that all content words can be classified exhaustively with the help of the two features [±nominal] and [±verbal]). In the following, the categories found at the terminal nodes of a phrase structure tree are called *terminal syntactic categories*. The term *syntactic category* refers to terminal and non-terminal (or phrasal) syntactic categories together. The term *lexical category* is only used in its second sense, which is defined in the following.

In its second usage, *lexical category* refers to the classification of lexical items according to grammatical criteria, i.e. the category information attached to each lexical entry in a dictionary. Although the category labels *noun*, *verb*, *adjective* and *preposition* are generally also part of this inventory, the category information given in a dictionary is usually much more fine-grained and comprises information concerning phonological, morphological and syntactic properties of the item in question.[10] In languages with complex inflectional morphology, for example, it is usually not sufficient simply to say that a given item is a noun or a verb. Instead, it has to be indicated that an item classified as a noun belongs to the second declension class, or that an item classified as a verb is intransitive and subject to an irregular passive formation. And so on.

It appears to be obvious that terminal syntactic categories are lexical categories in the sense that they are a necessary part of each lexical entry. That is, each lexical item has to be marked for a terminal syntactic category in order to be inserted at the right place into a phrase marker. No doubt, when stated in this very general way, it would be hard to disagree with this view. However, it is common to assume that terminal syntactic categories and lexical categories are commensurate in that lexical categories are but further subcategorisations of the more general terminal syntactic categories. That is, declension classes are but a further subcategorisation of the superclass of nouns, verb classes just a further subcategorisation of the superclass of verbs, etc. Such a neat correlation between terminal syntactic categories and lexical categories in fact appears to exist in a number of lan-

[10] Often, of course, there is also other information regarding, for example, pragmatic, sociolinguistic, or lexico-semantic properties of the lexical item. Here, however, we are only interested in the grammatical classification in a strict sense, i.e. the information needed to use a lexical item in a grammatically acceptable way.

guages (including, in particular, the Indo-European languages), but this is not universally so.

Consider the following hypothetical example: The most basic phrasal categories in language L always consist of an overt function word X and a content word Y, and practically all content words may co-occur with each function word. That is, the simplest and most straightforward way to describe the phrase structure of this language is this:

(23) $[_{XP}$ [X] [Y]]

where X represents function words, and Y content words. Phrases would then be distinguished primarily by the function words serving as the head of the phrase (i.e. determiner phrases consist of a determiner and a content word, auxiliary phrases of a tense/aspect auxiliary and a content word, and so on).[11] Consequently, all lexical items representing content words would simply be marked as Y (content word) in the lexicon. For the analysis of phrase structure, no further subcategorisation of the content words is required.

Assume further that language L allows the pluralisation of some but not all content words. Moreover, there are two very different plural marking strategies, one involving a suffix, the other one a prefix. That is, with regard to plural marking, the content words of language L fall into three classes: those which take plural suffixes, those which take plural prefixes and those which do not allow pluralisation. Semantically (ontologically), these classes roughly correspond to English count nouns, English verbs, and English mass nouns, respectively. In language L, then, it would be correct to claim that nouns and verbs are lexical categories but that these lexical categories are, strictly speaking, irrelevant for the analysis of phrase markers. In terms of phrase structure both nouns and verbs are simply content words.

While it may be relatively easy and straightforward to accept the claim that there may be a lack of correlation between terminal syntactic categories and lexical categories with regard to morphology, the correlation may also be lacking with regard to distributional, i.e. syntactic, criteria. This possibility arises, once again, simply because the lexical categorisation procedure tends to be much more fine-grained than the one concerned with terminal syntactic categories. Assume a language L1 in which the terminal syntactic categories of content words are indistinct in the same way as in language L above. Assume further that in L1 there are two function words marking

[11] Obviously, this is very similar to much recent work where it is assumed that functional categories are the heads of the overall construction.

negation, and that one of these two function words is consistently used with one class of content words and the other one with all of the remaining content words. That is, the negation words provide a syntactic environment which neatly classifies the class of content words into two classes which, based on ontological criteria, could be termed nouns and verbs. Again, L1 would be an example of a language where noun and verb are lexical categories but where this distinction plays no role with regard to terminal syntactic categories.

In the next section, it will be argued that Tagalog exhibits some of the properties of these hypothetical languages. To conclude this section, it may be noted that the proposed distinction of two different levels in the analysis of the grammatical properties of lexical items is similar in spirit and kind to the distinction of different levels (or tiers) in the analysis of clause structure made in a variety of non-transformational syntactic theories, including LFG. In line with these theories, the current proposal rejects the categorial uniformity hypothesis (Bresnan 1994:72) with regard to lexical items on two counts. First, it does away with the assumption that there is a simple universal grid according to which the lexical items of all languages can be grammatically classified. Second, it proposes that there are (at least) two distinct levels on which lexical items have to be grammatically analysed and categorised.

One level is the level of terminal syntactic categories where lexical items are categorised according to their phrase-structural properties (this level corresponds to the level of c(onstituent)-structure in LFG). The second level is the level of lexical categories proper where lexical items are categorised according to those grammatical features which are not directly relevant for phrase structure. This level pertains to a set of possibly very heterogeneous features (i.e. phonological, morphological and syntactic features). It is not unlikely that in a more detailed analysis it will turn out that rather than dealing with one level here it may be useful to make a further distinction between two or more levels on which this set of features can be adequately dealt with. For the purposes of this paper, however, it will be assumed that the two levels distinguished above are sufficient to account for all the relevant phenomena.

The distinction between two levels in the grammatical analysis of lexical items implies the possibility of different alignments between the two levels. In the most general terms, there are two possibilities: a) there is a correlation between the classes on both levels; b) there is no such correlation. The former possibility is well-known from Indo-European languages, both old and modern. The latter possibility is found in Tagalog (and possibly other Austronesian languages) and probably also in Salishan languages (see Jelinek and Demers (1994) for a discussion).

In both Tagalog and Salishan languages, however, the lack of a correlation between the two levels is due to the fact that content words are categorially indistinct with regard to syntactic categories. Logically, it would also be possible to find no correlation due to the fact that there is a distinct set of lexical categories and a distinct set of terminal syntactic categories in a given languages and that there is no correlation between the two sets. Furthermore, it is logically possible that the lexical items of a given language belong to distinct terminal syntactic categories but are indistinct with regard to lexical categories. Table 1 summarises these possibilities.[12]

Table 1: Possible alignments between lexical and syntactic categories

	LEXICAL CATEGORIES		TERMINAL SYNTACTIC CATEGORIES
Ia	distinct	≠	distinct
Ib	distinct	=	distinct
II	indistinct		distinct
III	distinct		indistinct
IV	indistinct		indistinct

Types Ia and Ib are distinguished by the fact that in Ia there is no correlation between lexical and syntactic categories while in Ib the two classes are commensurate.

This is not the place for a detailed discussion of whether all the possibilities provided for in Table 1 are actually attested in natural languages.[13] However, we may note that Type Ib is exemplified by Indo-European languages, Type III by Tagalog, Salishan languages and also Cayuga as analysed by Sasse (1993b). Type II may be attested among highly isolating languages. Several much-discussed agglutinating languages such as Turkic or Eskimo may be approaching Type Ia in that similar morphology is found across different syntactic categories and hence morphologically based classes may be orthogonal to classes based on phrase-structural positions. As for type IV, I am not aware of a language which looks like a promising candidate for the complete absence of any kind of grammatical distinctions among lexical items (despite the fact that much of the literature on categorial squishes tends to portray languages in such a way that a complete absence of grammatical distinctions is implied).

[12] See Sasse (1993b:200) for a very similar table. However, there is one crucial difference between Sasse's table and the one presented here. In Sasse's table *syntactic categories* refers to both terminal and non-terminal syntactic categories. Here, only the terminal ones are included.

[13] See Walter (1981), Sasse (1993a/b), Broschart (1997), and Anward et al. (1997) for further discussion and references to the fairly extensive literature on noun/verb squishes.

4 Lexical and Syntactic Categories in Tagalog

This section provides a more detailed discussion of the claim that in Tagalog there is no, or only a minimal, distinction between nouns and verbs. As already mentioned in the preceding section, this claim refers to the fact that Tagalog content words do not have to be subcategorised with regard to terminal syntactic categories. However, the way this claim is generally presented in the literature is somewhat confusing in that it is often implied that there are also no lexical categories in Tagalog. This second claim is wrong, as will be shown in section 4.2. But to begin with, in section 4.1 I will briefly repeat some of the arguments for the first claim, i.e. that Tagalog content words do not have to be categorised with regard to terminal syntactic categories in Tagalog.

4.1 The Lack of Terminal Syntactic Categories in Tagalog

Tagalog exemplifies the hypothetical languages discussed in section 3.2 with regard to two essential features. First, all phrasal categories, with the exception of the predicate (when clause-initial), are composed of a function word which indicates the category and a content word. That is, Tagalog clauses generally follow a very simple pattern in which there is a regular alternation between a single function word and a single content word:[14]

(24) *iniabót* ***ng*** *manggagamot* ***sa***
i-in-abót ng manggagamot sa
CV-RL(UNDR)-reach GEN doctor LOC
sundalo ***ang*** *itlóg*.
sundalo ang itlóg
soldier SPEC egg
'The physician handed the egg to the soldier.'

Second, content words are not subcategorised with regard to which function words they may co-occur with. Furthermore, all content words may also be used in predicate function. That is, almost all Tagalog content words may occur in exactly the same number and kinds of terminal positions in a phrase structure tree. Thus, assuming Kroeger's (1993:118-148) analysis of Tagalog clause structure, the structure of (24) can be analysed as follows (GP = genitive phrase, CW = content word):

(25) [IP [INFL [$_{CW}$ iniabót]] [S [GP [$_{GEN}$ ng] [$_{CW}$ manggagamot]] [PP [$_{P}$ sa] [$_{CW}$ sundalo]] [DP [$_{D}$ ang] [$_{CW}$ itlóg]]]]

[14] This basic pattern is usually somewhat obscured by the second-position clitics which occur in almost every Tagalog clause.

That almost all Tagalog content words may occur in exactly the same phrase-structural positions has been amply demonstrated in the literature quoted in footnote 5 above. Therefore, a few examples will suffice to support this claim.

One prominent feature of Tagalog phrase structure is the fact that not only OBJECT-denoting words may co-occur with the specific article *ang*. As illustrated by *tumulong* in *(4)* above and *áalagaan* in the following example, it is also possible (and not uncommon!) to use fully inflected (i.e. voice and mood-marked) ACTION-words in construction with the specific article:

(26)

iuuwi		*nya*	***ang***
i-RED1-uwí'		niyá	ang
CV-RED1-returned.home		3SG.POSS	SPEC
àalagaan	***nya***		
RED1-alaga'-an	niyá		
RED1-cared.for-LV	3SG.POSS		

'He would return the ones he was going to care for.'

Fully inflected ACTION-words also occur with the other markers of nominal expressions, i.e. *ng* and *sa*:

(27)

at	*ang*	*pare*	*at*	*siyá*	*ay*	*naghintáy*
at	ang	pare	at	siyá	ay	nag-hintáy
and	SPEC	priest	and	3SG	PM	RL.AV-wait
ng	***sàsabihin***		*ng*	*sundalo*		
ng	RED1-sabi-in		ng	sundalo		
GEN	RED1-statement-PV		GEN	soldier		

'And the priest and he waited for what the soldier would say.' (Bloomfield 1917:30/13)

(28)

nakàtanaw	*siyá*	*ng*	*bahay*	*na*
nakà-tanaw	siyá	ng	bahay	na
RL.STAT-in.sight	3SG	GEN	house	LNK
mailaw	***sa***	***pinatùtunguhan***		
ma-ilaw	sa	in-pa-RED1-tungo-an		
STAT-light	LOC	RL(UNDR)-??-RED1-direction-LV		
ng	*kalabáw*			
ng	kalabáw			
GEN	caribou			

'He saw a lighted house in the direction toward which the caribou was going.' (Bloomfield 1917:72/6)

Furthermore, fully inflected ACTION-words may be in construction with quantifiers, including the existential quantifier *may*:[15]

(29)

mayroon		*palang*	***nagàalaga***
may-doón		palá-ng	nag-RED1-alaga'
EXIST-DIST.LOC		so!-LNK	RL.AV-RED1-cared.for
doón	*sa*	*ibun*	
doón	sa	ibon	
DIST.LOC	LOC	bird	

'In fact, there was already someone looking after those birds.'

(30)

may	***ipàpakita***	*ako*	*sa*	*iyo*
may	i-RED1-pa -kita	akó	sa	iyo
EXIST	CV-RED1-CAUS-visible	1SG	LOC	2SG.DAT

'I have something to show you.'

In all of the preceding examples, one could insert content words of any other ontological or lexical category in place of the ACTION-words used here. Changing the category of the content word does not have any consequences whatsoever for the syntax and semantics of the overall phrase. For example, an existential quantifier phrase expresses existential quantification regardless of the ontological or lexical category of its complement: *may langgám* means 'there are ants' and *may ipàpakita* means 'there are things to be shown'. Similarly, the formal properties of the existential quantifier phrase do not vary with the kind of content word serving as the complement of the function word: there is never a linker between *may* and its complement and no clitics may intervene between these two constituents (but clitics may come in between *mayroón* and its complement and there is also a linker in the *mayroón* construction). A *may*-phrase may not only serve as the predicate of an existential or a possessive construction, but it may also serve as the complement in an *ang*-, *ng*- or *sa*-phrase. And so forth.

Consequently, it appears to be possible and useful to analyse the phrase structure of this and similar phrases simply as function word plus content word and to make no categorial distinctions between content words with regard to terminal syntactic categories. That is, all content words may occur, without further derivation or conversion, in the same kind of phrase-structural positions. In the following, this analysis will be referred to as the *syntactic uniformity hypothesis for content words*.

[15] In presentative constructions the existential quantifier is typically combined with a deictic, hence *mayroón* 'there is/was'.

There are a few alternatives to the syntactic uniformity hypothesis for content words. It is not uncommon to assume that in examples such as the preceding ones, the ACTION-words are somehow nominalised by the function words with which they occur, or that these examples involve headless relative clauses. While I am not aware of any irrefutable arguments which rule out such alternative analyses altogether, there are a number of arguments which render these analyses less plausible than the syntactic uniformity hypothesis for content words.

First, and perhaps most importantly, no formal differences whatsoever exist between ontologically different classes of words such as ACTION-words, PROPERTY-words, and OBJECT-words when occurring in the same phrase-structural position. Hence, if the occurrence of an ACTION-word in a determiner phrase or a quantifier phrase is interpreted as some kind of nominalisation (or a headless relative clause) there is no principled reason to exclude the same analysis for OBJECT-words (for example, *ang langgám* could be analysed as 'the one which is an ant', etc.). To generalise the alternative analyses in this way to all kinds of content words is of course just another way of stating the syntactic uniformity hypothesis for content words. To limit these analyses only to ACTION-words, on the other hand, is arbitrary unless independent evidence is adduced to show that the constructions with ACTION-words are indeed different from those containing other kinds of content words.

Second, the alternative analyses are less economical in that they posit an additional (and invisible) layer of structure (in the case of headless relative clauses) or additional morphological processes (in the case of nominalisations). The syntactic uniformity hypothesis for content words allows the most general and economical statement of the syntax and semantics of Tagalog phrase structure. Hence it is the preferred analysis for reasons of simplicity.

Third, from a crosslinguistic point of view, any analysis of Tagalog clause structure should be able to express the fact that ACTION-words can be used in the same phrase-structural positions as OBJECT-words without any extra morphological marking and the fact that such use is clearly more common in Tagalog than in languages which require nominalisation or relative clause formation in order to achieve the same functional effects. The syntactic uniformity hypothesis provides a straightforward account of this difference. It is unclear how the alternative analyses would be able to account for it.

4.2 Lexical Categories of Tagalog Roots

This section is concerned with the claim that Tagalog roots show different formal properties which are not directly predictable from the ontological

category of their denotata, and that in this sense they belong to different lexical categories. It thus presents a challenge to the view that Tagalog roots are precategorial, an issue to be taken up at the end of this section.

The major formal distinctions between Tagalog roots pertain to the affix sets with which they may occur. That is, the major parameter for the distinction of different lexical categories in Tagalog is a morphological one. To emphasize the morphological basis of the classification I will henceforth speak of the *morpho-lexical* classes of Tagalog roots.[16]

The fact that Tagalog roots belong to different morpho-lexical classes is not obvious (otherwise there would be no need to discuss this issue here). That is, there are no easily identifiable paradigms into which roots enter, nor is there anything like a set of distinct conjugation or declension classes as they are well-known from the older Indo-European languages. This is not to say that so far no morpho-lexical classifications of Tagalog roots have been proposed. Quite to the contrary, a substantial number of morpho-lexical classifications of Tagalog roots exists, most of them concerned with voice affixations.[17] However, the various classification proposals differ so widely that one wonders whether the authors are dealing with the same empirical domain. This should become obvious simply by looking at the number of classes proposed by different authors. Here is a fairly representative, but not comprehensive, list of the proposals found in the literature:

- Blake (1925:38f), who continues the work of the Spanish grammarians, operates with 17 classes of 'active verbs'.
- Schachter and Otanes (1972) work with the notion of an affix correspondence set, i.e. 'a set of two or more major affixes which, together with the base, form major transitive verbs of differing focus [= voice, NPH] but otherwise identical meaning' (1972:293). With this methodology they distinguish 33 classes for 'major transitive verbs' (1972:295-306) and 10 classes for 'major intransitive verbs' (1972:306-310), allowing the possibility that roots belong to more than one class. On the basis of the same methodology, Cruz (1975) proposes 38 classes while McFarland (1976:33) lists 47 'inflectional patterns for verbs', noting that 'most verb roots' occur in more than one pattern.

[16] This terminology is also motivated by the fact that it leaves open the possibility that further classifications of Tagalog roots are possible, based on other grammatical features (for example, phonological features) and not necessarily commensurate with the morpho-lexical classification sketched here.

[17] Work concerned with this topic is usually called 'the subcategorisation of Tagalog verbs' or something similar. However, since practically all Tagalog roots may occur with voice affixations, these classifications in fact propose a classification of all roots, not just ACTION roots.

- Ramos (1974, 1975), using Fillmorean 'deep cases', needs only 15 classes. De Guzman, who uses a similar methodology though with the additional assumptions of the Lexicase framework, distinguishes 7 major classes with 48 subclasses for 'primary verb stems' and an additional 14 major classes with 32 subclasses for 'secondary verb stems' (1978:243ff, 385ff).

The differences between these classificatory proposals are due in part to the different grammatical frameworks in which these authors work. What is more important, however, is the fact that all of these classifications make essential reference to the English translations of the Tagalog data, thus actually providing a cross-classification of Tagalog and English verbal expressions.

The major obstacle to an easy and straightforward morpho-lexical classification of Tagalog roots is the fact that there is pervasive polysemy (and possibly also homonymy) with regard to the affixes which may be used for classifying roots. That is, identifying a morpho-lexical class is not simply a matter of determining which roots occur with which formative since most formatives may occur with most roots. There is, for example, a highly productive prefix *ma-*, often glossed as a STATIVE marker, which may occur with the large majority of all Tagalog content word roots. However, it is possible and useful to distinguish two different kinds of formations involving *ma-*.[18] In one kind of formation *ma+ROOT* means 'have ROOT, be characterised by what the root denotes'. Let's call this the HAVE-formation. Examples include the following:

(31)	*mabahay*	'having many houses on it'	< *bahay* 'house'
	mabahá'	'flooded'	< *bahá'* 'flood'
	magandá	'beautiful'	< *gandá* 'beauty'
	madalí'	'quick'	< *dalí'* 'quickness'
	marami	'many'	< *dami* '(large) quantity, amount'[19]

In the second formation, the BECOME-formation, *ma+ROOT* means 'become ROOT, get into the state denoted by, or associated with, the root'.[20] For example:[21]

[18] For a more detailed discussion of *ma*-formations, proposing a similar but not identical analysis, see Wolff (1993).

[19] Intervocalic /d/ often alternates with /r/ in Tagalog.

(32)	*malutás*	'get solved'	< *lutás* 'solved'
	mahinóg	'become ripe, ripen'	< *hinóg* 'ripe'
	màmura	'become cheap'	< *mura* 'cheap'
	maduróg	'become crushed'	< *duróg* 'crushed, splintered'
	maluto'	'be/become cooked'	< *luto'* 'cooked, cuisine'
	maputol	'get cut off'	< *putol* 'a cut, a piece'
	mabutas	'get a hole, be perforated'	< *butas* 'a hole'
	matapos	'be/become completed'	< *tapos* 'end, conclusion'
	magalit	'become angry'	< *galit* 'anger'
	magutom	'become/feel hungry'	< *gutom* 'hunger'

The difference in meaning between these two formations is also reflected in formal differences. The BECOME-formation allows for aspectual and modal inflection. For example, apart from the basic form *magalit* there is also the realis perfective form *nagalit* 'became angry', the realis imperfective form *nagàgalit* 'was/is becoming angry' as well as the non-realis imperfective form *magàgalit* 'will become angry'. The HAVE-formation, on the other hand, generally does not allow any aspectual or modal inflection (there is no *nadalí' 'was quick', etc.). As opposed to BECOME-formations, however, HAVE-formations allow for simple (unstressed) reduplication to express plurality: *(mga) madadalí'* 'quick ones', *(mga) magagandá* 'beautiful ones', and so on.

Granted that there are two clearly different formations with *ma-*, these two formations provide the basis for two basic morpho-lexical classes of Tagalog roots: roots that occur in the HAVE-formation (class A), and roots that occur in the BECOME-formation (class B).[22] To be a useful morpho-lexical classification of Tagalog content word roots, this classification presupposes that class membership is disjunctive, i.e. that no root is a member

[20] Note that this formation includes both achievements and accomplishments and thus 'become' here has a wider meaning than the BECOME- operator used in Vendler-type lexical decomposition.

[21] In many instances, the BECOME-formation also has an abilitative interpretation. Thus, *malutás* also means 'can be solved', *maputol* 'can be cut off', etc.

[22] A small number of roots partake in neither formation. This is true in particular of roots referring to HUMAN BEINGS (e.g. *babae* 'woman, female', including KIN TERMS (e.g. *iná* 'mother'), with the exception of *anák* 'child, offspring' which allows the derivation of *maanák* 'having many children' (according to Santos (1983) *bata'* 'child, young' as well allows the derivation of *mabata'* 'having many children', but other dictionaries do not list this formation).

of more than one class. That this is in fact the case is supported by two further observations. First, a large number of the roots which occur in class B either denote a state all by themselves (such as the first five items in (32)) or allow the derivation of state-denoting expressions via stress shift. This holds true for the remaining five items in (32):

(33)

putol 'a cut, a piece'	*putól* 'be cut'
butas 'a hole'	*butás* 'perforated'
tapos 'end, conclusion'	*tapós* 'finished'
galit 'anger'	*galít* 'angry'
gutom 'hunger'	*gutóm* 'hungry'

That is, there is a good reason why members of class B generally do not allow the HAVE-formation: The meaning derived from roots via HAVE-formation for class A roots is either inherent in class B roots or can be achieved through a different process.

The second observation pertains to the fact that it is possible to derive accomplishment readings from class A roots via a different formation. This is done by infixing *-um-* into those class A roots for which an accomplishment reading is semantically feasible:

(34)

gandá 'beauty'	*gumandá* 'become beautiful'
dalî 'quickness'	*dumalî* 'become fast'
dami '(large) quantity, amount'	*dumami* 'become many'

These two observations make it clear that roots from both classes allow the derivation of the same kinds of meanings but that they employ different formal means in the process (see also Table 2). Therefore, the difference between the two classes is a *grammatical* difference. It is not due to the ontology of the entities denoted by class A and B roots, respectively (i.e. it is not the case that the formation *ma-*+CLASS A ROOT means 'have ROOT, be characterised by what the root denotes' simply because all class A roots are THING roots, and that the formation *ma-*+CLASS B ROOT means 'become ROOT, get into the state denoted by, or associated with, the root' simply because all class B roots are PROPERTY roots).

Table 2: Distribution of form and meaning in expressions for states and accomplishments based on class A and B roots

	class A	class B
STATE	ma-	0/stress shift
ACCOMPLISHMENT	-um-	ma-

In fact, although there is some kind of correlation between ontological categories and the two morpho-lexical categories just established, this correlation is far from perfect. Roughly speaking, it holds true that most roots denoting THINGS, ANIMALS and NATURAL PHENOMENA (floods, earthquakes, etc. as well as mountains, rivers, etc.) belong to class A while those denoting STATES, PROCESSES, and ACTIONS belong to class B. However, PROPERTY roots are split between classes A and B. For example, 'beauty', 'quickness', 'quantity, plentitude' are class A (see (31)) while 'ripeness', 'cheapness', 'anger' and 'hunger' are class B (see (32)).

A further morpho-lexical subclassification of the two classes just established appears to be possible, in particular with regard to the formation of actor voice forms by affixing either *-um-* or *mag-*. Without going into details, class A roots generally allow the formation of actor voice forms only by prefixing *mag-* (e.g. *magbahay* 'build one's own house'). Many class A roots, however, do not allow the derivation of an actor voice form directly from the root (**magbahá'*, **maggandá*, **magdalî'*). Class B roots, on the other hand, generally allow the formation of an actor voice form directly from the root. For the large majority of class B roots this is possible by infixing *-um-* (e.g. *lumutás* 'to solve, to clear up)', others only allow *mag-* (e.g. *magluto'* 'to cook'), while a third class allows both *-um-* and *mag-* with a difference in meaning (e.g. *pumutol* 'to cut off' vs. *magputol* 'to cut into pieces/several things' or 'to cut oneself').

However, as in the case of the two *ma-*formations, the subclasses involved here are far from obvious. To provide a sound basis for these subclasses requires a lengthy discussion of various polysemies, exceptions and overlaps, a task well beyond the limits of this paper. Here the task was simply to establish the fact that morpho-lexical classes exist in Tagalog. And the establishment of the two major disjunctive classes based on the different *ma-*formations should be sufficient to make this point.

If one accepts the claim that Tagalog roots belong to different morpho-lexical classes, it follows that Tagalog roots are not precategorial. Or, to put this a bit more carefully, at least two possible interpretations of the term

precategorial are clearly not applicable to Tagalog.[23] In one interpretation—the one introduced by Verhaar (1984:2)—this term refers to bound roots (i.e. roots that do not occur unaffixed) from which items belonging to different lexical or syntactic categories (nouns and verbs, for example) can be derived, without there being clear evidence that one of the possible derivations of a given root is more basic than the other one. No significant subset of Tagalog roots exists which could be characterised as precategorial in this sense for the simple reason that there are no bound roots. This will be demonstrated in detail for ACTION roots in the following section. The examples in section 3.1 amply illustrate the fact that it is also not the case for OBJECT roots (cf. *langgám*, *bata'*, *asawa*, *bahay*, etc. in examples *(1)-(22)*). One example for a PROPERTY root is the following one:

(35)	*tingnan*	*mo*	*ang*	***gandá***	*na*	*ng*	
	tingin-an	mo	ang	gandá	na	ng	
	look-LV	2SG.POSS	SPEC	beauty	now	GEN	
	buhay	*ni*	*Maria*				
	buhay	ni	Maria				
	life	GEN.PN	Mary				

'Look how beautiful Maria's life is now!'
(lit. look at the beauty of Maria's life now!)

In a second interpretation, *precategorial* may refer to the fact that roots, though not necessarily bound, are categorially indistinct with regard to grammatical features. That is, all kinds of derivations are possible from all kinds of roots. The enormous productivity and polysemy of Tagalog formatives may give rise to the impression that this is indeed the case. However, as shown in this section, it is a misconception: Tagalog roots belong to different morpho-lexical classes and hence are categorially distinct and do not allow just any derivation that is semantically appropriate.

Tagalog roots may be deemed to be categorially indistinct with regard to terminal syntactic categories. But, as argued in section 4.1, this feature they share with all Tagalog content words, roots as well as derived words. And to call Tagalog roots precategorial with regard to terminal syntactic categories is something of a misnomer since there is no later (derived) categorial stage with which the *pre*categorial stage could be contrasted.

[23] See also Clynes' (1995:203-205) reservations about the usefulness of this concept for the description and analysis of Austronesian languages.

5 Uses and Meanings of Tagalog Action Roots

Almost all Tagalog roots can be used, and *are* used, without further affixation. Of particular concern here are those roots which denote ACTIONS, and the present section will deal only with these roots. When ACTION roots are used without further affixation, they may convey one of the following kinds of meaning: (a) the state which ensues from the successful performance of the action (similar to a past participle in English); (b) the result[24] or the typical or cognate object of the action (similar to object(ive) nominalisations in English); or (c) the name of the action (similar to an action nominalisation in English). As the following examples will show, meanings (a) and (b) are found primarily with roots denoting transitive (or ditransitive) ACTIONS, while meaning (c) is found with roots denoting transitive and intransitive ACTIONS.

Many roots convey more than one of the three meanings just mentioned. In fact, a few roots may convey all three of them. Furthermore, it is not always possible clearly to distinguish between the different kinds of meanings, in particular between meanings (a) and (b) as well as between meanings (b) and (c). Still, the semantic and syntactic context generally provides enough clues to determine which meaning is intended in a given example.

This section, then, makes two points: First, it empirically supports the claim that Tagalog ACTION roots can be used without further affixation. This point is proved by all of the well-formed examples in this section. Furthermore, at the end of the section I will quote some data from the corpus study by McFarland (1976) which shows that unaffixed use is not only a grammatical possibility but actually occurs in moderate frequency in natural data and thus clearly is a fact of everyday language use.

The second point is the claim that unaffixed roots convey distinctly different kinds of meanings. In order to show this, it is necessary to define contexts (test frames) in which only roots with one kind of meaning can occur while the others cannot. Throughout the discussion of these test frames it is important to keep in mind that what is tested for is semantic well-formedness. If roots conveying a particular meaning cannot be used in context X this is because their meaning is not compatible with the meaning of the other items in the construction. Grammatically, roots behave like all other content words. That is, in principle they fit into every position open to content words. Hence the difference in meaning between roots conveying

[24] Note that 'result' here refers to a THING, i.e. the THING which results from the action. Thus, for example, the result of cutting here is 'a piece (cut off)' and not the state of being cut. See also the comment below after example *(51)*.

meaning (a) and roots conveying meaning (b) should not be misconstrued as a difference in syntactic distribution (for example, that only roots conveying meaning (a) can occur in predicate position while those conveying meaning (b) cannot).

Roots denoting states can be distinguished from roots denoting the result, object or name of an action by the fact that only the state-denoting roots can be used as modifiers, as in the following two examples:

(36) *ang* ***lutás*** *na* *problema*
ang lutás na problema
SPEC solved LNK problem
'the solved problem'

(37) *ang* ***nakaw*** *na* *kabayo*
ang nakaw na kabayo
SPEC stolen LNK horse
'the stolen horse'

Roots denoting the result, object or name of an action cannot occur in this function. Thus, for example, the root *putol* 'a cut, a piece cut off' cannot be used in the following phrase to express the indicated meaning:

(38) *#ang* ***putol*** *na* *kalye*
SPEC cut LNK street
'the cut off (i.e. blind) street'

The symbol # is used here to indicate that a given structure is ill-formed with regard to the indicated meaning. As just noted, these examples are not ungrammatical since they conform to the basic morphosyntactic rules of Tagalog. As shown in section 4.2, Tagalog content words (including roots) are generally not subcategorised for a specific syntactic function. That is, a root such as *putol* may, in principle, occur in any syntactic function, provided that it makes sense within the overall construction. Thus, *putol* may occur in a linker construction, provided that it does not have to be interpreted as a modifier. Compare the following example:

(39) *ang* *bakíl-bakíl* ***na*** *putol* *(ng* *buhók*
SPEC uneven LNK cut GEN hair
ni *Huán)*
PN.GEN John
'(John's) uneven (hair) cut'

Note also that *putol na kalye* in *(38)* could be given an appositional meaning (??'the cut off piece, the street') but that is in all likelihood not something one would ever say in this way.

The fact that *putol* cannot be used as a modifier is not the only evidence for the claim that this root denotes the result of the ACTION of cutting ('a cut, a piece cut off') rather than a state ('be cut off'). Another piece of evidence consists of the fact that there is an overt derivation from the same root which denotes a state, i.e. *putól* with stress on the ultimate syllable which means 'cut off, severed'. The stress-shifted *putól* can easily function as a modifier. Compare *(38)* with:

(40) *ang* ***putól*** *na* *kalye*
SPEC cut LNK street
'the cut off (i.e. blind) street'

The change in meaning brought about by the different stress pattern is also shown by the fact that *putól* can no longer co-occur with another modifying element:

(41) *#ang* *bakíl-bakíl* ***na*** ***putól*** *(ng* *buhók*
SPEC uneven LNK cut GEN hair
ni *Huán)*
PN.GEN John
'(John's) uneven (hair) cut'
(a possible but non-sensical interpretation would be ??'the uneven one cut off')

There is a substantial number of ACTION roots in Tagalog which are similar to *putol* in that they denote the result or object of the action when stressed on the penult, but a state when stressed on the ultima. Examples include: *butas* 'hole'—*butás* 'perforated', *tapos* 'end, conclusion'—*tapós* 'finished, done', *bali'* 'a break, fracture'—*bali'* 'broken', *bayad* 'payment, fee, charge'—*bayád* 'paid', etc. (see also (33) and Schachter and Otanes 1972:196f, Wolff et al. 1991:374f). That is, the two different kinds of meanings attributed here to different roots are not only relevant to explaining the differences in distribution between two types of ACTION roots. The

same distinction has also to be made in order to explain the two different stress patterns shown by a substantial number of ACTION roots.

Turning now to roots denoting the name of an action, there are two kinds of test frames for these roots. First, these roots may be used as the subject of predicates which denote the manner in which an event/action took place. For example:

(42) *biglaan* *ang* *kanyáng* ***alís***.
biglá'-an ang kanyá-ng alís
sudden-?? SPEC 3SG.DAT-LNK departure
'His departure (act of leaving) was sudden.' (English 1986)

(43) *Subali't* *tuluy-tulóy* *pa* *rin* *ang* ***kain***
subalit RED5-tulóy pa din ang kain
but RED5-continue still also SPEC eating
ni *Matsíng*.
ni matsíng
PN.POSS small.monkey
'But the monkey's eating continued nevertheless.'
(Wolff et al. 1991:526)

Roots denoting a state or the result/object of an action may not occur in this context since the resulting clauses would be semantically ill-formed (??'her stolen one was sudden', ??'his load was sudden').

Second, roots denoting the name of an action cannot be used as predicates in clauses with personal pronouns as subjects:

(44) *#alís siyá.* 'S/he left/is leaving' (??'s/he is a departure')

(45) *#kain akó* 'I am eating/ate/was eaten'
(??'I am consumption of food')

This constraint also holds for roots denoting results/objects:

(46) *#putol siyá* 's/he was cut/has cut sth' (??'s/he is a cut')

But it does not hold for roots denoting states:

(47) *nakaw siyá* 's/he was/is stolen'

summarises the diagnostic contexts used here to distinguish the three different meanings which can be conveyed by ACTION roots.

Table 3: Diagnostic contexts for distinguishing the different meanings of ACTION roots

	modifier	subject of manner predicate	predicate with pronominal subject
root denotes state	YES	NO	YES
root denotes result or typical object of action	NO	NO	NO
root denotes name of action	NO	YES	NO

As already mentioned in the introduction to this section, it is not the case that all ACTION roots convey only one of the three meanings listed in . Some roots have both a result/object and a state meaning. Examples include *ayos* which means 'order, arrangement' as well as 'presentable, fit to be seen', *bigáy* 'gift' and 'given', *bagsák* 'a (sudden) fall' and 'failed, defeated', *dalá* 'load, cargo' and 'carried',[25] *gawá*' 'an act, product' and 'made, caused by', *hango*' 'relief, extract, removal' and 'extracted, derived', etc.[26] The following two examples illustrate the two meanings of *dalá*:

(48) *ang* ***dala*** *naming* *bala*
ang dalá namin.LNK bala
SPEC carried 1PL.EX.POSS.LNK bullet
'The bullets carried by us...'

(49) Inilapág ni Hwán ang
i-in-lapág ni Hwan ang
CV-RL(UNDR)-space.below PN.POSS John SPEC
kanyang **dalá**
kanyá-ng dalá
3SG.DAT-LNK load
'Juan dropped his burden...' (Bloomfield 1917:106/16)

[25] Interestingly, the Spanish loan *karga* (and its variant *kargá*) conveys the same meanings as the native form *dalá*, at least according to English (1986) and Santos (1983). According to Panganiban (1972) *karga* is 'load' and *kargá* is 'carried'.

[26] As can be seen from this rather short list, many—but not all!—roots in this class are stressed on the final syllable. Hence stress shift cannot be used to differentiate the two kinds of meaning.

And here are two examples for the two meanings of *bigáy*:

(50) *P350 every two months ang **bigáy***
P350 ang bigáy
P350 SPEC given
ko-ng bayad sa school.
ko-ng bayad sa
1SG.POSS-LNK payment LOC
'The payment given by me to the school is 350 Pesos every two months.'

(51) *Parang **bigáy** na lang natin*
para-ng bigáy na lamang natin
as.though-LNK gift now only 1PL.IN.POSS
sa kanilá yon.
sa kanilá iyón
LOC 3PL.DAT DIST
'That is like our gift to them.' (i.e. consider it a gift to them)
(Wolff et al. 1991:1050)

The overlap between the state and the result/object meanings should not come as a surprise since even on an ontological level the difference between a THING and a PROPERTY/STATE is not always easy to discern (cf. also Sasse 1993b:202). Thus, the THING 'gift' may also be viewed as something which is in the STATE of being given, i.e. 'the given one/thing', the THING 'load, cargo' is 'the carried one', and so forth.

Similarly, polysemy involving the two meanings 'result/object of an action' and 'name of an action' is found in many languages, in particular with regard to nominalisations (cf. Koptevskaja-Tamm 1988:42). English *filling*, for example, may refer to the actual act or manner of filling (as in *With completion of filling, net-like anastomoses were noted*) and hence be an action nominalisation (the name of an action). Or it may refer to the cognate object (as in *manifold fillings were prepared*) and hence be interpreted as an objective nominalisation. Not surprisingly, then, many Tagalog ACTION roots also allow both readings, as illustrated by *lakad* 'walk' in the following examples:

(52) *Yaon ay mahabang **lakad.***
iyón ay ma-haba'-ng lakad
DIST PM STAT-length-LNK walk
'That was/is a long walk.' (English 1986)

(53) *Mahusay ang **lakad** ng mákiná.*
ma-husay ang lakad ng mákiná
STAT-orderliness SPEC walk GEN machine
'The walking (running condition) of the machine is good.'
(English 1986)

Thus, there is not only an overlap between roots denoting results/objects and those denoting states, but also one between roots denoting results/objects and those denoting the name of an action.

In passing, it may be noted that for each kind of meaning there is a typical set of syntactic functions in which it most commonly occurs. For instance, roots denoting states typically occur in predicate or modifier function. See examples *(36)* and *(37)* above and the following example:

(54) *inilúgaw na, **lutó** na*
i -in-lugaw na luto' na
CV-RL(UNDR)-gruel now cooked now
'It (the rice) has already been made into porridge, it is already cooked.'

Roots denoting the result or object of an action typically occur in nominal expressions:

(55) *ang sakít ng mga **hampás** niyá*
ang sakít ng mga hampás niyá
SPEC pain GEN PL blow 3SG.POSS
sa akin balikat!
sa aking balikat
LOC 1SG.DAT.LNK shoulder
'How painful the blows he gave me on my shoulders!' (English 1986)

Roots denoting the name of an action are also typically found in nominal expressions, in particular those functioning as the subject of manner predicates:

(56) *at habang bumìbilis ang* ***takbó***
at habang um-RED1-bilís ang takbo
and while INGR-RED1-speed SPEC run
ng tubig
ng tubig
GEN water
'And when the current of the water got faster...'

However, the fact that there are typical contexts of use for each type of meaning does not mean that the meaning(s) conveyed by a given root are simply a product of the context of use. That is, it is not the case that a root such as *takbó* means 'act of running, a run' simply because it is typically used in nominal expressions functioning as the subject of manner predicates. If the meaning of a root depended simply on its context of use, it could not be explained why some uses of a root are impossible (for example, why it is not possible to say *#ang takbóng baboy* 'the running pig'). Furthermore, that the meaning conveyed by a given root is independent from the syntactic function in which it is used is also shown by the fact that identical meanings are conveyed by bare roots in at least two different syntactic contexts. For example, roots denoting the name of an action do not exclusively occur in nominal expressions. Instead, use in predicate position is also possible:

(57) ***iyák*** *ang sagót niyá sa akin*
iyák ang sagót niyá sa akin
cry SPEC answer 3SG.POSS LOC 1SG.DAT
'His answer to me was crying/to cry.' (elicited)

(58) ***datíng*** *niyá ang hindí ko*
datíng niyá ang hindí' ko
arrival 3SG.POSS SPEC NEG 1SG.POSS
alam
alam
knowledge
'I don't know (about) her arrival.' (elicited)

To conclude this discussion of the meaning of bare ACTION roots, it has to be pointed out that some unaffixed ACTION roots also occur in a variety of uses and functions in which they do not convey the three meanings discussed so far. The data on these further uses and functions are somewhat limited and it is highly likely that they can be shown to be derived rather

than basic uses and functions.[27] Perhaps the most prominent of these further uses is the use in imperatives illustrated by the following examples:

(59) ***Hampás*** *na kayó, mga bata', sa mga*
hampás na kayó mangá bata' sa mangá
blow now 2PL PL child LOC PL
langgám.
langgám
ant
'Whip at the ants, boys.' (Bloomfield 1917:221/42)

(60) *Umuwí' na tayo, Daddy!* ***Uwí'***
um-uwí' na tayo uwí'
AV-returned.home now 1PL.IN returned.home
na tayo!
na tayo
now 1PL.IN
'Let's go home Daddy! Let's go home!'

As clearly shown in particular by the second example, unaffixed roots used as imperatives correspond to actor voice forms (i.e. the subject is the addressee of the command). Note that the standard form of imperatives in Tagalog involves voice-marked forms and that actor voice as well as undergoer voice imperatives are possible (cf. Schachter and Otanes (1972:402-409). Obviously, the imperative meaning arises here from the overall construction (and the situational/textual context) and is not in any way directly linked to the bare roots functioning as predicates. Furthermore, according to the native speakers I have consulted (though generally not noted in the literature), the ability to use a bare root as an imperative is limited to those roots which form their actor voice with the infix *-um-* (cf. Himmelmann 1987:166). Thus, it is not a general characteristic of ACTION roots.

It may also be noted that, according to some dictionaries, a small number of ACTION roots never occur without affixes. Examples from English (1986) include roots such as *agaw* (with voice affixes: 'to snatch, to grab'), *patol* (with voice affixes: 'to notice, to pay attention'), *suno'* (with voice affixes: 'to give someone a lift'), *ubos* (with voice affixes: 'to consume, to use up'). Rubino (1998a) gives a nominal translation for *ubos*, Panganiban

[27] For more extensive discussion, see Bloomfield (1917:218-223), Himmelmann (1987:157-171) and Wolff et al. (1991:115, 291, 488, 1028f, 1130-1134).

(1972) gives one for *agaw* and *suno'* (but not for *ubos*), while Santos (1983) gives one for all four roots.

These differences in lexicographic practice should, in my view, be interpreted as reflecting the fact that ACTION roots differ substantially with regard to the frequency and naturalness of unaffixed uses. For a number of roots, including the three roots just mentioned, such use is probably highly unusual and hence not recorded in some of the dictionaries. This, however, does not mean that it is completely impossible to use such roots without affixes, given an appropriate context.

Most importantly, it should be clearly understood that in general unaffixed uses of ACTION roots are not in any way exceptional. Instead, they are reasonably common both in terms of types (the number of roots for which such use is attested in natural data) as well as in terms of tokens (the number of times a bare ACTION root can be found in a corpus). McFarland (1976), for example, has found unaffixed uses in naturally occurring (written) texts for 76 of the 106 most common ACTION roots in his corpus (= ca. 70%). For some roots such as *dalá* 'load; carried', unaffixed use is in fact the single most frequent use (i.e. it is more frequent than the use of any voice-marked form). In short, although there are differences with regard to how frequent and natural it is for a given ACTION root to occur without affixes it is clearly the case that Tagalog ACTION roots quite generally allow for unaffixed uses.

6 Root Meaning and the Derivational Nature of Voice Marking

It follows from the data presented in the preceding sections that voice marking in Tagalog is derivational in all its manifestations. While the derivational nature of voice marking is widely accepted in the case of OBJECT roots, it is controversial with regard to ACTION roots (see de Guzman (1997) for a recent survey of this controversy). This section presents an explication of the derivational nature of Tagalog voice marking, based on the data discussed so far.

To begin with, we may note that there is a widespread, though usually implicit, assumption that Tagalog voice marking is derivational in at least some of its uses. In all approaches which make a distinction between nouns and verbs as syntactic or lexical categories, the following assumption has to be made: when affixed to 'nominal' roots (i.e. roots denoting THINGS or ANIMATE BEINGS) voice marking has two functions. First, it derives a verb from a noun, and second, it registers the alignment between semantic role and syntactic function with regard to the subject argument. For example,

adding *i-* to *anák* to form *ianák* 'to give birth to someone' converts the putative noun *anák* 'child' into a verb 'to give birth'. Furthermore, the prefix *i-* registers the fact that the subject argument is a displaced theme. Other examples of this kind include the following:

(61)

bahay 'house'	*magbahay* 'to build one's own house'
anák 'child'	*mag-anák* 'to breed, to have a child'
buladór 'kite, rocket'	*magbuladór* 'to fly a kite'
walís 'broom'	*magwalís* 'to sweep'

Thus, voice marking appears to have derivational force in at least some of its manifestations.[28] Here, however, the much more general claim is made that voice marking is derivational in *all* its manifestations. The discussion will be limited to the potentially most controversial issue, i.e. the use of voice marking morphology on ACTION roots (i.e. those roots which in many analyses are considered verbal roots). It is claimed that, as in the case of OBJECT roots, voice marking on ACTION roots also has a dual function. First, however, we have briefly to establish the fact that *all* voice-marked words in Tagalog, regardless of their base, are members of a single morpho-lexical class and that this morpho-lexical class is different from all other morpho-lexical classes. This class is called here the 'V-class' and its members 'V-words'.[29]

The major formal criterion for V-class membership is voice marking: All and only words bearing voice affixes are members of this class. The fact that voice-marked words constitute a special morpho-lexical class is shown by a number of further formal and semantic properties shown only by the members of this class. Formally, only V-words are inflected for aspect and

[28] An alternative analysis for these examples, which does not attribute derivational (category-changing) force to voice marking, would involve the assumption of zero conversion along the following lines: The nominal root $anák_1$ 'child' is converted into a verbal root $anák_2$ 'to give birth' which then is inflected for voice to register different alignments between semantic role and syntactic function. A major weakness of this alternative analysis is the fact that there is no independent evidence for a zero-converted verbal root $anák_2$ 'to give birth'. That is, under no circumstances does the form *anák* without further affixation mean 'give birth'.

[29] The 'V' here stands for 'voice-marked'. In principle, one could also call the members of this class 'verbs' as long as it is clearly understood that V-words are here defined as a morpho-lexical category. That is, 'V' does not indicate a syntactic category and therefore is not fully commensurable with the most common and best established use of the term *verb*, i.e. as the name of a syntactic category. Furthermore, having a class of verbs in a given language is usually taken to imply that there is also a class of nouns in that language. To date I have considerable doubt as to whether this implication would be true in Tagalog. That is, it is far from clear whether it is possible and useful to identify a morpho-lexical category 'noun' in Tagalog. See also section 10.5.

mood. In fact, aspect-mood marking and voice marking are formally inseparable. Hence, each voice-marked form is also aspect-mood-marked, as shown by the aspect-mood paradigm in Table 4:

Table 4: Aspect-mood paradigm for *bilí* 'buy'
(r = realis, n-r = non-realis)

	act	PV	LV	CV
n-r perfective	b-um-ilí	bilh-ín	bilh-án	i-bilí
n-r imperfective	bìbilí	bìbilh-ín	bìbilh-án	i-bìbilí
r perfective	b-um-ilí	b-in-ilí	b-in-ilh-án	i-b-in-ilí
r imperfective	b-um-ìbilí	b-in-ìbilí	b-in-ìbilh-án	i-b-in-ìbilí

The forms found in the first row of Table 4 are generally considered the unmarked forms with regard to aspect-mood marking (Schachter and Otanes (1972:66) call them *basic forms*). These are used in imperatives, in control constructions and as non-initial predicates in clause chains. Note that aspect-mood inflection is fully predictable (unlike voice marking). From each basic form the other three forms are derivable via totally general and nearly exceptionless rules (cf. Schachter and Otanes 1972:361-371).

Aspect-mood inflection is not the only formal characteristic of V-words. Further morphological characteristics include the fact that a number of other derivations are predictable on the basis of the actor voice form. For example, gerunds are formed by prefixing *pag-* to the roots which take *-um-* as the actor voice affix (e.g. *pumutol* 'to cut off'—*pagputol* 'cutting off') while roots taking *mag-* as the actor voice prefix form gerunds by prefixing *pag-* + unstressed reduplication of the root or stem-initial syllable (e.g. *magluto'* 'to cook'—*pagluluto'* 'cooking'). For further discussion and references see Schachter (1996:44-46).

Semantically, V-words differ from all other content words in Tagalog in that they are systematically ambiguous. That is, all V-words have two readings depending on the syntactic context. Used as predicates, they denote a specific instance of the action denoted by the root, as in:

(62) ***sumigáw*** *yung* *anák*
um-sigáw iyón.LNK anák
AV-shout DIST.LNK child
'That child shouted:...'

Used as modifiers in a linker construction, they have the same reading:

(63) *yung* *anák* *na* ***sumisigáw***
iyón.LNK anák na um-RED1-sigáw
DIST.LNK child LNK AV-RED1-shout
'that shouting child'

Used in nominal expressions, however, V-words denote one of the participants involved in the ACTION denoted by the root:

(64) *yung* *anák* *ang* ***sumigáw***
iyón.LNK anák ang um-sigáw
DIST.LNK child SPEC AV-shout
'The one who shouted/the shouter is (was) that child.'

Note that in this use it is also a specific instance of the action denoted by the root that is being referred to by the voice-marked form *sumigáw*. That is, *sumigáw* may mean 'the one who shouted/the shouter on this particular occasion'. It cannot mean 'a professional shouter/someone who always shouts'.[30]

Having thus established the fact that V-words belong to a special morpho-lexical class of their own, it follows almost automatically that voice marking is derivational. At least, it is category-changing in that it derives V-words from non-V-words. Still, if ACTION roots generally were bound roots it could be argued that a substantial difference existed between the affixation of voice affixes to OBJECT roots and the affixation of voice affixes to ACTION roots. Only in the case of OBJECT roots voice marking would involve a category shift while in the case of ACTION roots it could be argued that the function of voice marking was simply to register the alignment between semantic role and syntactic function with regard to the subject argument.

However, as demonstrated in sections 3.2 and 5, ACTION roots are not some kind of bound forms which do not have a specific meaning and cannot be used by themselves. Instead, in their unaffixed use they denote states and/or results/objects or names of an action. Hence, adding a voice affix to an ACTION root does not simply register a change in alignment between semantic role and syntactic function. It derives a different lexical item: A root denoting a state or the result of an action is turned into an *actor-* or *undergoer-oriented action* expression. More specifically, this derivation

[30] In this regard, Tagalog V-words differ from various kinds of oriented nominalisations in English. English *shouter* is ambiguous between a specific instance reading ('the one who shouted just now') and a non-specific type reading ('someone who always/professionally shouts').

involves two aspects. First, the derived form denotes the actual and specific performance of an action. Second, the specific performance of an action is not denoted in some neutral way but rather the voice-marked form is oriented towards one of the participants: an actor voice form denotes the agent of the action, hence in its predicative use the subject has to be an agent. A patient voice form denotes the patient of the action and in its predicative use requires the subject to be a patient. And so on. There is no substantial difference here between OBJECT and ACTION denoting roots. Both aspects are present whenever voice marking is added to any kind of root, including OBJECT ROOTS and ACTION roots.

Apart from accounting for the facts discussed in this paper, this derivational view of voice marking in Tagalog is also supported by a number of phenomena which remain unexplained (or which are hard to explain) in other accounts. For example, the derivational view provides an easy and straightforward explanation for the ubiquity of voice marking in Tagalog. Most roots (and a large number of derived stems) allow the affixation of at least one voice affix without any further derivation. If voice marking were not in itself derivational and could only be attached directly to verbs, this ubiquity could only be accounted for by postulating pervasive homonymy on the root level or by assuming extensive use of zero conversion. There is no independent evidence to support these alternative analyses.

Moreover, the proposed view also provides a straightforward explanation for the pervasive formal and semantic idiosyncrasies of Tagalog voice marking, which are typical of derivational processes. Here the following simple examples will suffice to illustrate these idiosyncrasies:[31] formally, there are unpredictable idiosyncrasies such as the deletion of root vowels in undergoer voice forms (e.g. the vowel /i/ is deleted in the patient voice form *kanin* 'to eat something' which derives from *kain* 'consumption of food'; the regular form *kainin* is also possible). Other examples of unpredictable formal idiosyncrasies are the sporadic insertion of /n/ in patient or locative voice forms such as *tawanan* 'to laugh at someone' (< *tawa* 'laugh, laughter, laughing') and completely irregular forms such as the patient voice form *kunin* 'to get something' (< *kuha* 'getting, a helping'), etc. (see also Bloomfield 1917:213-215).

Semantically, the meaning of V-words is broadly predictable on the basis of the meaning of the root and the meaning of the voice affixes. Thus, for example, it is clear that the subject of an actor voice form is an agent of some kind and that the action denoted by the voice-marked form is in some

[31] For further discussion and examples see, among others, Himmelmann 1987:129-146 and Rubino 1998b:1152-1155.

way related to the THING or ACTION denoted by the root. But there are manifold idiosyncrasies with regard to both the notion of agency and the notion of action conveyed by an actor voice form. Thus, *mag-anák* (from *anák* 'child') does not mean 'to give birth'. Instead, it means 'to breed, have a child' and is most commonly used of animals (the actor voice form for 'to give birth' is *manganák*). The root *talo* means 'defeated, surpassed, beaten' but the derived form *magtalo* does not mean 'to defeat, to beat' (that is the meaning of *tumalo*). Instead, *magtalo* means 'to contend, to quarrel, to disagree'. Similarly, *magtaká*, which is derived from *taká* 'surprise, surprised', does not mean 'to surprise someone' but 'to be surprised'.

In short, there is a considerable number of facts which strongly suggest that voice marking in Tagalog is derivational. This in itself is not an original observation (see de Guzman (1997) and Rubino (1998b) for references). The present account, however, differs from that found in the literature in that it provides a more precise and explicit statement of what this derivation involves, in particular with regard to ACTION roots. This is achieved by providing a testable explication of the different meanings conveyed by roots and a definition of the meaning and the morpho-lexical class of V-words.

7 Conclusion

An important aspect of the voice system in Tagalog is the fact that voice-marked words (V-words), as well as the roots from which they are derived, belong to lexical categories which are very different from the lexical categories found in more familiar languages such as English. It has been repeatedly suggested that the difference pertains to the fact that Tagalog roots are precategorial and/or that there is no distinction between nouns and verbs in Tagalog. The preceding sections present a somewhat different explication of the difference between the two systems of lexical categories and its repercussions for the voice systems. Specifically, it is claimed that Tagalog roots are generally not bound and/or precategorial roots. Instead, Tagalog roots belong to different morpho-lexical classes. That is, it cannot be predicted solely on the basis of their meaning with which affixes a given root may occur. Furthermore, all kinds of roots, including roots denoting ACTIONS, allow for unaffixed uses. In their unaffixed uses, roots may denote THINGS, ANIMATE BEINGS, PROPERTIES, STATES, RESULTS OF ACTIONS, NAMES OF ACTIONS, etc., i.e. concepts which in English are generally rendered by nouns or adjectives. What roots cannot denote is the actual and specific performance of an action. Only V-words may denote the actual performance of an action. In this regard, Tagalog V-words are similar to English verbs. They differ, however, from English verbs in a number of important respects outlined below.

While Tagalog V-words clearly belong to their own morpho-lexical category, they do not belong to a special terminal syntactic category. That is, V-words have morphological and semantic properties which set them apart from all other Tagalog content words. But with regard to the positions they may occupy in a phrase structure tree, they do not differ from other content words. In English, on the other hand, there is an unambiguous correlation between (morpho-)lexical and terminal syntactic categories: Membership in a given lexical (sub-)category implies membership in a specific terminal syntactic category ('die' is an intransitive verb, and hence a verb and not a noun).

All Tagalog V-words are necessarily derived, while in English there are both basic and derived verbs. The derivation of Tagalog V-words is manifest both formally and semantically: formally, the morpho-lexical category of the root is changed to the morpho-lexical category of V-words. Semantically, an oriented action expression is derived from an expression which denotes a THING, STATE, NAME OF AN ACTION, RESULT OF AN ACTION, etc. The notion *oriented action expression* conveys two things: first, oriented action expressions denote the actual performance of an action (and not the name or the result of an action). Second, they denote the actual performance of an action in such a way that at the same time they also denote one of the participants involved in the action.

For English verbs active voice is the basic, non-derived voice. For Tagalog V-words all voices are derived in the same way. Hence, it does not make sense to consider one of the four Tagalog voices the basic voice.

8 References

Anward, Jan, Edith Moravcsik & Leon Stassen. 1997. Parts of speech: A challenge for typology. *Linguistic Typology* 1:167–183.

Artawa, Ketut and Barry J. Blake. 1997. Patient Primacy in Balinese. *Studies in Language* 21:483–508.

Blake, Frank R. 1925. *A Grammar of the Tagalog Language*. New Haven: American Oriental Society.

Bloomfield, Leonard. 1917. *Tagalog Texts with Grammatical Analysis*. Urbana, Illinois: University of Illinois. Three volumes.

Bresnan, Joan. 1994. Locative inversion and the architecture of Universal Grammar. *Language* 70:72–131.

Bresnan, Joan W. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

Broschart, Jürgen. 1997. Why Tongan does it differently: Categorial distinctions in a language without nouns and verbs. *Linguistic Typology* 1:123–165.

Capell, Arthur. 1964. Verbal systems in Philippine languages. *Philippine Journal of Science* 93:231–249.

Clynes, Adrian. 1995. *Topics in the phonology and morphosyntax of Balinese, based on the dialect of Singaraja, North Bali*. Ph.D thesis, Australian National University, Canberra.

Cruz, Emilita L. 1975. *A Subcategorization of Tagalog Verbs*. Quezon City: University of the Philippines. *The Archive* Special Monograph No. 2.

de Guzman, Videa P. 1978. *Syntactic Derivation of Tagalog Verbs*. Honululu: University Press of Hawai'i.

de Guzman, Videa P. 1997. Verbal affixes in Tagalog: Inflection or derivation?. In *Proceedings of the Seventh International Conference on Austronesian Linguistics* ed. Cecilia Odé & Wim Stokhof. 303–325. Amsterdam: Rodopi.

DeWolf, Charles M. 1979. *Sentential Predicates: A Cross-Linguistic Analysis*. Ph.D thesis, University of Hawai'I, Honululu.

DeWolf, Charles M. 1988. Voice in Austronesian languages of Philippine type: passive, ergative, or neither?. In *Passive and Voice* ed. Shibatani, Masayoshi. 1988. 143–193. Amsterdam: Benjamins.

English, Leo J. 1986. *Tagalog-English Dictionary*. Manila: National Book Store.

Gil, David. 1993. Tagalog Semantics. *BLS* 19:390–403.

Himmelmann, Nikolaus P. 1987. *Morphosyntax und Morphologie—Die Ausrichtungsaffixe im Tagalog*. München: Fink.

Himmelmann, Nikolaus P. 1991. *The Philippine Challenge to Universal Grammar*. Arbeitspapier Nr. 15. Köln: Institut für Sprachwissenschaft.

Himmelmann, Nikolaus P. 1997. *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur*. Tübingen: Niemeyer (Linguistische Arbeiten 362).

Himmelmann, Nikolaus P. 1998. Regularity in irregularity: Article use in adpositional phrases. *Linguistic Typology* 2:315–353.

Jackendoff, Ray. 1983. *Semantics and Cognition*. Cambridge, Massachusetts: MIT Press.

Jelinek, Eloise & Richard A. Demers. 1994. Predicates and pronominal arguments in Straits Salish. *Language* 70:697–736.

Keenan, Edward L. 1976. Towards a Universal Definition of 'Subject'. In *Subject and Topic* ed. Charles Li. 305–333. New York: Academic Press.

Koptevskaja-Tamm, Maria. 1988. *A typology of action nominal constructions*. Ph.D thesis, Stockholm University.

Kroeger, Paul R. 1993. *Phrase Structure and Grammatical Relations in Tagalog*. Stanford: Stanford University Press.

Lemaréchal, Alain. 1982. Semantisme des parties du discours et semantisme des relations. *Bulletin de la Société de Linguistique de Paris* 77:1–39.

Lemaréchal, Alain. 1989. *Les parties du discours: Sémantique et syntaxe*. Paris: P.U.F.

Li, Charles N. (ed.). 1976. *Subject and Topic*. New York: Academic Press.

Lopez, Cecilio. 1977. Preliminary Study of Affixes in Tagalog. In *Selected Writings in Philippine Linguistics,* Cecilio Lopez. 28–104. Quezon City: University of the Philippines. First published 1937.

McFarland, Curtis D. 1976. *A Provisional Classification of Tagalog Verbs.* Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Müller, Friedrich. 1882. *Grundriss der Sprachwissenschaft.* Bd. II, Abt. 2. Wien: Alfred Hölder.

Naylor, Paz B. 1995. Subject, Topic, and Tagalog syntax. In *Subject, Voice and Ergativity* ed. David Benett, Theodora Bynon & George B. Hewitt. 161–201. London: SOAS.

Panganiban, José V. 1972. *Diksyunario-Tesauro Pilipino-Ingles.* Quezon City: Manlapaz Publishing Co.

Pittman, Richard, 1966, Tagalog *-um-* and *mag-*. An Interim Report. *Papers in Philippine Linguistics* 1:9–20. Canberra: Pacific Linguistics. (A-8).

Ramos, Teresita V. 1974. *The Case system of Tagalog verbs.* Canberra: Pacific Linguistics (Series B-27).

Ramos, Teresita V. 1975. The Role of Verbal Features in the Subcategorization of Tagalog Verbs. *Philippine Journal of Linguistics* 6:1–24.

Rubino, Carl R. G. 1998a. *Tagalog Standard Dictionary.* New York: Hippocrene Books.

Rubino, Carl R. G. 1998b. The morphological realization and production of a non-prototypical morpheme: the Tagalog derivational clitic. *Linguistics* 36:1147–1166.

Santos, Vito C. 1983. *Pilipino–English Dictionary.* 2nd revised edition. Metro Manila: National.

Sasse, Hans-Jürgen. 1993a. Syntactic Categories and subcategories. In *Syntax* ed. Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann. 1993. 646–686. Berlin: de Gruyter.

Sasse, Hans-Jürgen. 1993b. Das Nomen—eine universale Kategorie?. *Sprachtypologie und Universalienforschung* 46:187–221.

Schachter, Paul. 1976. The Subject in Philippine Languages: Topic, Actor, Actor-Topic or None of the Above. In *Subject and Topic* ed. Charles Li. 491–518. New York: Academic Press.

Schachter, Paul. 1996. *The Subject in Tagalog: Still none of the above.* UCLA Occasional Papers in Linguistics No.15. Stanford: UCLA.

Schachter, Paul & Fay Otanes. 1972. *Tagalog Reference Grammar.* Berkeley: University of California Press.

Scheerer, Otto. 1924. On the Essential Difference Between the Verbs of the European and the Philippine Languages. *Philippine Journal of Education* 7:1–10.

Shibatani, Masayoshi (ed.). 1988. *Passive and Voice.* Amsterdam: Benjamins.

Verhaar, John W. M. 1984. Affixation in contemporary Indonesian. In *Towards a description of contemporary Indonesian: Preliminary Studies, Part I,* ed. Bambang Kaswanti Purwo. 1–26. Jakarta: Universitas Atma Jaya. NUSA 18.

Walter, Heribert. 1981. *Studien zur Nomen-Verb-Distinktionaus typologischer Sicht.* München: Fink.

Wolff, John U. 1993. Why roots add the affixes with which they occur. In *Topics in Descriptive Austronesian Linguistics* ed. G. P. Reesink. 217–244. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië. *Semaian* 11.

Wolff, John U. with Maria Theresa C. Centeno and Der-Hwa V. Rau. 1991. *Pilipino through Self-Instruction.* Ithaca: Cornell Southeast Asia Program. Four volumes.