J. Linguistics **36** (2000), 215–259. Printed in the United Kingdom
© 2000 Cambridge University Press

Formalizing formality: an analysis of register variation in Sinhala[1]

JOHN C. PAOLILLO

University of Texas at Arlington

(Received 3 March 1998; revised 6 August 1999)

Variation in language on the basis of formality (register variation) is often neglected both in grammatical descriptions and in sociolinguistic analyses. I demonstrate here that in Sinhala, and perhaps in other diglossic languages, register variation in syntax cannot be ignored. In a Head-Driven Phrase Structure Grammar (HPSG) analysis based on a corpus of naturally occurring Sinhala texts, I propose an analysis of register variation in which the syntax of all observed registers is accounted for within a single grammar. I further explain how the approach to register variation developed here can be extended to other types of sociolinguistic variation.

1. Introduction

Register variation – variation in the form of linguistic expressions according to the formality of the social context of use – is a widely recognized type of variation in language, and it is sometimes used to frame theoretical arguments in syntactic analyses. Yet register occupies no consistent place in syntactic analysis generally, and the considerable register-based syntactic variation in individual languages has yet to receive an adequate theoretical treatment. As an illustration of this problem, consider the treatment of pied-piping in *wh*-movement constructions in English since Ross (1967). Pied-piping variants of *wh*-movement constructions, as in (2), have a stylistic effect that is decidedly more formal than the non-piped variants (1).

(1) Reports which the government prescribes the height of the lettering on the covers of are a shocking waste of funds.
(2) Reports the height of the lettering on the covers of which the government prescribes are a shocking waste of funds. (Ross 1967: 109)

[1] This work began as my dissertation research, supported in part by a grant from the National Science Foundation (BNS-9013361). I would specifically like to acknowledge the following people for insightful comments and discussion on previous versions of this paper: Emily Bender, Chris Culy, Michael Darnell, Jim Gair, Susan Herring, Ivan Sag, Peter Sells, Elizabeth Traugott, Rev. K. Nāgita, Rev. T. Sutadhara and two anonymous but thoughtful *JL* referees. Any errors of fact or analysis are the sole responsibility of the author.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

Even though this stylistic effect is as robust as any grammaticality judgment, and even though it is often noted in passing, it has never been systematically treated as part of the analysis of pied-piping. Instead, analyses tend to focus exclusively on the parameters of syntactic variation, leaving aside the question of how (1) and (2) come to reflect different registers of English.

Occasionally, a researcher may choose to exploit the stylistic markedness of pied-piping constructions for other analytic purposes. For example, consider the observations about register in the comparison of pied piping in English and German offered in Webelhuth (1992: 118, 128–132). Webelhuth uses the 'bookishness' of pied-piping in English relative clauses to argue that examples like (2) are not true relative clauses and so should not be handled by the same grammatical mechanisms. He further suggests that the stylistic value is a consequence of the use of a relative pronoun as an 'indexical' pronoun (132). However, in a discussion of *wh*-question pied-piping in German (118), a set of examples specialized for 'very colloquial style' receives no special treatment, and no explanation is offered for its stylistic value. Apart from these two cases, register variation is otherwise un-noted in Webelhuth's comparison of Germanic syntax.

The problems that arise when register is not taken into account are even more acute in languages with a great degree of register variation. At its most extreme, register variation takes the form of diglossia (Ferguson 1959, Hudson 1992), in which a language has two principal registers, High and Low, that are structurally divergent. In such situations, formal, written communications typically use grammatical structures associated only with H, while informal, spoken communications use corresponding structures associated only with L. Many well-studied languages are diglossic – including Arabic, Javanese, Tamil, Swiss German and Sinhala – and many others, if not diglossic at present, have diglossic periods in their history – including Bengali,[2] Chinese, French and other Romance languages, Hindi, Hebrew, Greek, Japanese and Telugu.[3] Yet observations about register are characteristically minimal in grammatical descriptions of most of these languages, even though register variables can differ in ways that are important to linguistic theory; for example, Arabic has different basic word orders for H (VSO) and many varieties of L (SVO), and Sinhala has markedly different morphosyntax in H and L (see below). Many descriptions of diglossic languages also note mixed or hybrid varieties of H and L which resemble codeswitching between distinct languages. In spite of this, and in spite of the current interest in syntactic constraints on codeswitching (Myers-Scotton

[2] Most of the major South Asian languages are historically or presently diglossic.

[3] For discussion of these and other diglossic situations, see A. Hudson (1991) and the references of individual papers therein. For Bengali, see Chatterji (1962, 1986). For Hindi, see King (1994), and for Telugu see Krishnamurti (1977). See also A. Hudson (1992) for a more detailed list of references on diglossic situations.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

1993, Milroy & Muysken 1995), register-switching in diglossic languages has received scant attention from a syntactic point of view.

The ultimate effect of neglecting register variation in syntax is that we are left with a limited understanding of the relation between the two. If register variation could be approached in a systematic way, however, then syntactic analyses could provide broader and more principled coverage of the relevant data. The present study begins to address this need through a grammatical analysis of register variation in Sinhala, a diglossic Indo-Aryan language spoken in Sri Lanka. With this analysis, I aim to illustrate three main points: (i) analysis of register variation is a necessity for an adequate grammatical description of Sinhala, (ii) no new formal apparatus is necessary for the analysis of register variation, and (iii) parallel analyses can be offered for codeswitching and other kinds of sociolinguistic variation, permitting the formalization of a more general theory of syntactic variation.

The remainder of this paper is organized as follows. Section 2 defines the notion of register used in this study, situating it with reference to other recognized uses of the term, and to related types of language variation. Section 3 describes some of the principal variables that distinguish Sinhala High and Low registers, and summarizes Gair's (1992) account of this co-variation in terms of Government-Binding theory. Section 4 presents evidence bearing on the distribution of the register variables from an analysis of a corpus of naturally occurring texts, where they are found to co-occur in ways not predicted by Gair (1992). A formal analysis accounting for these new observations framed in Head-driven Phrase Structure Grammar (HPSG) is presented in section 5, and a summary and conclusions are presented in section 6.

2. Register variation and diglossia

Terms such as 'register' are used in various and conflicting ways, so linguists are often faced with the problem of interpreting inconsistent definitions when making observations about language use. An apt illustration of this difficulty is encountered in the analysis of cooking recipes in English (Haegeman 1987, Culy 1996). A number of linguists concur that 'recipe English' constitutes a register of English, including Quirk, Greenbaum, Leech & Svartvik (1985),[4] Biber (1988, 1994) and Ferguson (1994). Yet these same linguists disagree on the nature of register. For Quirk et al., registers are determined by considerations of discourse topic ('field of discourse'), whereas for Ferguson, they are determined by the whole communicative situation (participants, setting, communicative function, etc.), and for Biber, any situational

[4] Although the definition of Quirk et al. (1985) is the one cited by Haegeman (1987), she indicates that she is investigating the register of 'cookbooks and other instruction manuals' (nearly all her examples are from recipes), or 'instructional English', which suggests something much broader, perhaps the procedural text-type of Longacre (1996).

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

parameter by itself can define a 'register'. Still others define the notion of register as occupationally determined varieties of language (Trudgill 1974, Wardhaugh 1992).

In the present study, I use the term register to denote a dimension of language variation that is orthogonal to dialect, genre, text type, modality and discourse topic. Under this view, the categories RECIPE and COOKBOOK are genres (Hymes 1972), or recurrent categories of message types, not registers. Individual recipes could be said to exemplify procedural discourse, a basic TEXT TYPE (Longacre 1996), contrasting with narrative, expository and hortatory text types. Recipes may use a written or spoken MODALITY, but are most characteristically written, and their DISCOURSE TOPICS are confined to the domain of cookery. A register, on the other hand, is at least in principle independent of these parameters. For example, within the foreigner talk register (Ferguson 1971), it is possible to have recipes, personal life stories, arguments, conversations and telephone messages (all genres); to have narratives, procedures, expositions, and other text types; to use either written or spoken modalities; and to discuss topics of academic, personal, medical, financial or any other concern. In all cases, the communication would still be recognizable as 'foreigner talk'. The same is true of academese, baby talk and colloquial English, all registers according to this definition. What sets registers in this sense apart from variation according to the other parameters above is that register variation involves, on the one hand, socially significant role relationships for the participants of the speech situation (such as teacher and student, official and citizen, parent and child, peer and peer), and on the other, an attitudinal characterization of the message itself (such as authoritative or ritually significant for formal registers, intimate or affectionate for baby talk registers, ratifying and accepting for peer-group informal registers, etc.).

Diglossia, in the sense of Ferguson (1959), is a type of register variation in which there are two principal structurally divergent registers of a language.[5] One register, the Low or L variety, is the colloquial variety that members of the speech community acquire in the course of natural development. It is the language of the home and family, and of casual everyday conversation. The other register, the High or H variety, is a formal variety that is learned only through formal instruction. It is generally the language of a valued and prestigious literature which is deeply connected with the culture of the speech community, and is used on formal occasions and for most written communication. The structural differences between H and L vary from case to case, but generally speaking, differences in phonology, lexicon, syntax and morphology characterize H and L. Use of one register in a circumstance

[5] In other words, excluding the extended cases of diglossia recognized by Fishman (1967) in which genetically unrelated languages are used for different functions in a society, such as Spanish and Guaraní in Paraguay. Much research on diglossia has also addressed the question of binarity, such as Britto (1986), Fasold (1984: 44–50) and Paolillo (1997).

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

calling for the other constitutes a serious faux pas that can engender discomfort (about the situation in which it arises) or laughter (usually at the expense of the speaker).

Three characteristics of diglossic languages point out the necessity of accounting for register in grammatical description. First, H and L are sharply differentiated in function, and the co-occurrence restrictions of H and L forms are strong enough that most occasions of use appear to demand H or L forms exclusively. On account of this, many researchers choose to write separate grammatical descriptions for H and L, as if they were separate languages. Second, H and L have clear structural relations in diglossic languages. This characteristic has led some researchers to propose parametric accounts of H and L differences (for example, Gair 1992, for Sinhala), in the spirit of the account of Germanic syntax by Webelhuth (1992), or that of Belfast English by Henry (1995). Third, 'mixed' registers are often noted in descriptions of diglossic languages, in circumstances that resemble conversational codeswitching between different languages (Gumperz 1982).[6] On the one hand, such mixed registers raise the possibility that an indeterminate number of grammars would have to be written to produce a complete grammatical account of a diglossic language. On the other hand, their existence permits parametric accounts of register variation to be empirically tested, by providing additional related varieties of language to examine.

Thus, in diglossic languages at least, register variation can hardly be ignored in grammatical description. Assuming that each separate grammatical description carries with it the cost of potential duplication of large sections of the grammar, we should hope to be able to account for the register variation of diglossic languages with a minimum number of separate grammars. Furthermore, the H and L registers of Sinhala are widely known by speakers of the language, and thus should be considered a single linguistic competence. In the account of Sinhala register variation I propose in section 5, a single grammar describes both H and L forms; this same grammar also accounts for the intermediate registers that are found in Sinhala.

3. Register variation in Sinhala diglossia

Sinhala diglossia has been well studied from the perspective of the grammatical differences between H and L (De Silva 1967, 1974, 1976; Gair 1968, 1986, 1992; Paolillo 1991, 1992, 1997).[7] A wide range of grammatical

[6] Ferguson comments that 'communicative tensions which arise in the diglossia situation may be resolved by the use of relatively uncodified, unstable, intermediate forms of the language' (Ferguson 1959 [1996: 31]). Among the earlier studied cases of conversational codeswitching is that of Norwegian in Hemnesberget (Blom & Gumperz 1972; Gumperz 1982: 60), where the switching is between what are described as diglossic levels of the language, Ranamål (L) and Bokmål (H).

[7] Although Paolillo (1991) is originally a later paper, and a criticism of Gair (1992), it appeared in print earlier through an accident of publishing.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

features is identified in these studies as distinguishing Literary Sinhala (H) and Colloquial Sinhala (L). Gair (1968, 1986) also identifies a third register, 'Formal Spoken Sinhala', used for public speaking purposes, but treats it together with Colloquial as L, since it appears to have Colloquial grammar with Literary lexical items giving it its formal flavor.[8] In this paper, I will focus on a set of features which Gair (1992) suggests are in a parametric relationship. These are subject-verb agreement, accusative case, the default case rule called the 'accusative rule', and the existence of a copula verb.[9] In addition to these features, I will also examine the distribution of locative case, as an example of a register variable that falls outside of Gair's parametric account. In the subsections below, I first describe each of the five register variables mentioned above, and then summarize the parametric account of Gair (1992).

3.1 *Agreement*

Perhaps the single, most significant indicator distinguishing Literary and Colloquial Sinhala is subject-verb agreement (Gair 1968: 10). Agreement in Literary involves person (first, second or third), number (singular or plural) and gender (masculine, feminine or inanimate). A sample past tense verb paradigm for *wæṭe-* 'fall', showing the full range of agreement categories, is given in Table 1 (masculine and feminine gender are not distinguished in the

		Singular	Plural
First		wæṭunemi	wæṭunemu
Second		wæṭunehi	wæṭunaahu
Third	masc.	wæṭuneeya	wæṭunooya
	fem.	wæṭunaaya	wæṭunooya
	inan.	wæṭuneeya	

Table 1
Agreeing forms of *wæṭe-* 'fall' (past tense)

present tense). In Colloquial, there is simply a single, invariant form *wæṭuna* 'fell' which serves as the past for subjects of all person and number categories. Likewise, in the present tense, Colloquial has the invariant form *wæṭenawa* 'fall' for all persons and numbers, while Literary has agreement for person, number and gender.

[8] Gair (1968) uses the term 'Spoken Sinhala' as a cover term for Colloquial and Formal Spoken.

[9] Gair (1992) also discusses the passive voice and other constructions which will not be treated here.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

Examples of contrasting sentences from naturally occurring texts, with and without subject-verb agreement, are given in examples (3) and (4) respectively. In (3), from a text in Literary, the third person plural form *giyoo* '(they) went' occurs, while in (4), oral narrative in Colloquial, the non-agreeing finite form *giyaa* 'went' occurs.

(3) Literary
owuhu bæňda damaa siṭi kumaariya laňgaṭa
they.NOM bind.PPL put.PPL exist.PPL.ADJ princess.ACC near.DAT
giyoo ya.
go.PST.3PL PRED
'They went near to the princess who had been bound.'

(4) Colloquial
itiṅ ee daruwa ehaa gedeṭṭa giyaa.
so that child.NOM there house.DAT go.PST.FIN
'So that child went to the house over there.'

3.2 *Copula verb*

Another characteristic that distinguishes Literary and Colloquial Sinhala is the distribution of the copula verb. In equational sentences, such as example (5), Literary uses a copula verb *we-* which, unlike other verbs with two direct arguments, takes both its arguments (subject and predicate) in the nominative case. Normally in Literary, one of the two direct arguments of a verb, namely the object, would be accusative.

(5) Literary
æyagee nama kamalaa winii.
her name Kamala COP.PST.3SG.FEM
'Her name was Kamala.'

The verb *we-* has other functions in Literary, namely as a copula in cleft sentences (6), and as an existential predicate (7). The functions of *we-* in examples (6) and (7) are absent in Colloquial.

(6) Literary
keeyas keḷee ṭenisange pæhæragæniimak no wee.
K.S. do.PST.EMPH Tennyson.GEN eliminate.NML.IND NEG COP.3SG
'It was not an elimination of Tennyson that K. S. did.'

(7) Literary
nawa nagara saṅkiini anikut goḍanægili atara
new city.GEN clustered other-also buildings among
nawa mahajana bæṅku shaakhaawada weyi.
new People's Bank branch.IND-also COP.3SG
'Among (the) other buildings clustered in the new city is a new People's Bank branch.'

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

Colloquial Sinhala, in contrast, forms equational sentences merely by juxtaposing two nominative case noun phrases, without any copula verb or other copular morphology, as illustrated in example (8). Sentences such as (8) are syntactically and semantically verbless; their time reference may be either present or past, as determined by the context, and there is no lexical verb in Colloquial that can be inserted in (8) which preserves the copular meaning. Furthermore, non-finite clauses in Sinhala are typically based on non-finite inflected verb forms, and without a verb to bear the inflection, sentences like (8) have no corresponding non-finite forms (Gair & Paolillo 1988). In contrast, the copula verb *we-* functions as a true verb in Literary, and has the full range of finite and non-finite forms, one example of which appears in (12). Thus, expressions based on the copula verb are possible in Literary which have no analog in Colloquial (for example, equational relative clauses: 'the one who is a teacher').

(8) Colloquial
rajjuruwoo diga ræwulak æti minihek.
king.NOM long beard.IND.NOM having man.IND.NOM
'The king was a long-bearded man.'

Existential sentences in Colloquial do not use *we-* but instead use either *in-* (for animate subjects) or *tiye-* (for inanimate subjects), both of which may be used in Literary.

The verb stem *we-* does appear in Colloquial, in the full complement of finite and non-finite forms. However, it always occurs with the process-verb sense 'become', rather than the copular meaning, as indicated in (9). The process meaning of *we-* is also found in Literary.

(9) Colloquial
taatta mærunayiṅ passe, dæṅ mee maaligaawe adipatiyaa
father die.PST.INST after now this palace.GEN master
unee, ee aananda kiyapu siṭuwareyaagee putaa.
we.PST.EMPH that Ananda called millionaire.GEN son
'After the father died, it was the son of the millionaire Ananda who now became (/*was) the master of this estate.'

For these reasons, I will assume in the analysis below that there are two distinct but homophonous verb stems *we-*, one being the copula, which is not present in Colloquial, and the other meaning 'become', which is present in both Literary and Colloquial. Since only the copular sense of *we-* is distributed asymmetrically over the different registers of Sinhala, I will not consider the 'become' sense further. The analysis in sections 4 and 5, however, distinguishes finite and non-finite copular forms of *we-*.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

3.3 *Accusative case*

Another grammatical feature that specializes for Sinhala Literary register is the accusative case. Accusative is only realized as a distinct case form in certain classes of nouns. These fall primarily into three types, all of which are animate: (i) personal pronouns, (ii) indefinite animate nominals, and (iii) plural animate nominals. Singular definite animate nominals generally do not exhibit distinct nominative and accusative case forms, with the exception of some feminine nouns (Gair & Karunatilaka 1974). Noun classes that do not have distinct nominative and accusative case forms (animate singular and inanimate nouns) use the nominative case form for all accusative functions. Examples of contrasting nominative and accusative case forms appear in Table 2.

NOM	ACC	gloss
mama	maa	1SG pronoun
api	apa	1PL pronoun
minihek	minihaku/miniheku	'man(INDEF)'
minissu/minisu	minisun	'people'
kumaari	kumariya	'princess'
rajjuruwo	rajjuruwan	'king'

Table 2
Nominative and accusative case forms

3.4 *Nominative and accusative default case* (*the* '*accusative rule*')

The interest of the accusative case in Sinhala register variation lies not only in the contrast in form between Literary and Colloquial Sinhala, but also in the distribution of the cases. In Colloquial, nominative forms serve both nominative and accusative functions.[10] Thus, nominative case functions as the 'default' case in Colloquial. In Literary, however, where nominative and accusative case are distinct both morphologically and syntactically, there is a strict co-occurrence relationship between nominative case and verb agreement. Verb agreement is absent in most subordinate clause types, and overt subjects of such clauses are obligatorily marked accusative. Moreover, nominative case is prohibited on objects of verbs and postpositions (including those taking nominative case in Colloquial). Thus, in Literary, accusative

[10] For this reason, the nominative case is sometimes referred to as the 'direct' case in descriptions of Colloquial Sinhala, since it is used to mark all direct arguments of the verb, as well as arguments of postpositions when another case (dative, genitive or instrumental) is not present. In this paper, however, I will retain the term 'nominative'.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

case functions as the default case for nominals in non-agreeing positions, a pattern referred to as the 'accusative rule' (Gair & Karunatilaka 1974). An example of accusative as the default case occurs in (10), where the modal adjective predicate *puḷuwan* 'can' has an infinitival non-agreeing clause complement, whose subject *owun* 'they (ACC)' and object *rajjuruwan* 'king (ACC)' are both in the accusative.

(10) Literary
owun rajjuruwan maranṭa puḷuwan ya yi,
they.ACC king.ACC kill.INF can PRED QUOT,
koolitaṭa hituṇee ya.
Kolita.DAT realize.PST.3SG PRED
'They could kill the king, Kolita realized.'

Both arguments would appear in the nominative in the corresponding Colloquial version of the sentence (*owuhu*[11] and *rajjuruwo*, respectively). A sentence from an oral narrative illustrating this appears in (11), in which the phrase *tamange seewakayo tuṅ hatara denek* 'three or four of his servants' functions as object of the verb *yæwwa* 'sent'.

(11) Colloquial
(mahawsada panḍitayoo) maaligaawaṭa gihilla, tamange
(Mahawsada pundit.PL.NOM) palace.DAT go.PPL self.GEN
seewakayo tuṅ hatara denek yæwwa.
servants.NOM three four ANIM.IND.NOM send.PST.FIN
'Having gone to his palace, Mahawsada sent three or four of his servants.'

The corresponding Literary sentence would require accusative case on both the noun 'servant' and the indefinite numeral 'three or four (animate indefinite)': *tamange seewakayan tuṅ hatara deneku*. Likewise, in Literary, the finite predication with agreement corresponding to (10) would appear with the subject in the nominative.

There is an important interaction of the accusative rule with the copula verb in Literary. If the copula verb is in a non-agreeing form, both subject and predicate are in the accusative case, as in example (12), where the embedded copula *weṇṭa* appears in the infinitive, a non-finite and non-agreeing form, subcategorized for in this construction by the modal *oonæ* 'must'. Thus, there is a general pattern of case agreement between subject and predicate in copular clauses, in which the nominative or accusative case of the subject is shared with the predicate. There is no direct analog of this

[11] Since Colloquial also uses different pronominal forms, the third person plural nominative pronoun that would actually be used in Colloquial in place of *owuhu* 'they (NOM)' would be either *eyaala* or *egolla*, both of which lack a morphological distinction between nominative and accusative case.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

behavior in Colloquial, where copula-less equational sentences like (8) always simply appear with both nominals in the nominative case.

(12) Literary
ohu horaku wenṭa oonææ yayi ohu duṭu
he.ACC thief.ACC COP.INF must PRD.QUOT he.ACC see.PST.ADJ
hæṭiyee koolitaṭa hituneeya
way.GEN Kolita.DAT think.INVOL.PST.3SG
'From the way he looked, Kolita thought, "He must be a thief".'

To summarize, there are two distinct and mutually exclusive default case systems in Sinhala. In one system, nominative case functions as the default case for subjects, equational predicates, and objects of verbs and postpositions. In the other system, nominative appears only in clauses with agreement, always on the subject, but also on the predicate if the verb is the copula *we-*. The behavior of these two distinct default case systems needs to be addressed in an account of Sinhala register variation.

3.5 *Locative case*

A final contrast in the case system present in Literary that is absent in Colloquial is the locative case, ending in *-hi* (see Table 3). In both Literary

LOC	GEN	gloss
kælææwehi	kælææwe	'in/of the forest'
awasthaawehi	awasthaawe	'in/of the opportunity'
widyaalayehi	widyaalaye	'in/of the school'
widyaalayakhi	widyaalayaka	'in/of a school'
snakhaarayanhi	sankhaarayanwala	'in/of the aggregates'

Table 3
Locative and genitive case forms

and Colloquial, locative and genitive case functions are not systematically distinguished, so the term 'locative' to describe *-hi* is potentially confusing. In Colloquial, there is only a single form, the 'genitive', normally ending in *-a*, *-e*, or *-ee* (and sometimes identical to the nominative case) for inanimates, and in *-gee* for animates. In Literary, both the locative and genitive forms appear, though the locative forms tend not to occur with semantically animate nouns.[12]

[12] This is the last vestige of the historically locative character of the locative case in Literary.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

Feature	H registers (Literary)	L registers (Colloquial)
finite forms	agreeing (2)	non-agreeing (3)
case system	nominative/accusative locative/genitive	syncretized (direct case) syncretized (genitive case)
default case	accusative (9), (11)	nominative (= direct) (10)
copula	copula obligatory (4)	no copula ever (7)

Table 4
Contrasting variables in Sinhala registers

Feature	H registers (Literary)	L registers (Colloquial)
agreement	+	−
accusative case	+	−
accusative rule	+	−
copula	+	−
oblique subjects	−	+
passive	+	−
generic Ø reference (pro_{arb})	−	+
AGR	+	−

Table 5
Literary and colloquial register variables and AGR

3.6 *A parametric account of Sinhala register variation*

A summary of the contrasting register variables of Colloquial and Literary Sinhala discussed above appears in Table 4, along with the numbers of the examples which illustrate them.

Working within Government-Binding theory (Chomsky 1981), Gair (1992) proposes a parametric account to explain the Literary and Colloquial co-occurrence patterns in Table 4. Other syntacticians have proposed similarly parametric accounts of variation, generally across different dialects of a language (for example, Webelhuth 1992, Henry 1995 and contributions to Black & Motapanyane 1996, especially Henry 1996). In Gair's analysis, a single parameter, the presence or absence of abstract agreement (AGR), accounts for the observations in Table 4. This is schematized as in the top half of Table 5.

According to Gair (1992), when AGR is present, it is realized as subject-

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

verb agreement,[13] governing the subject position and assigning it nominative case; furthermore, nominative is reserved for noun phrases governed by AGR. This necessitates that objects of verbs and postpositions will appear in a different (default) case, the accusative.[14] The copula is necessitated as a morphological host for AGR in equational sentences. Gair (1992) notes additional properties of Literary and Colloquial, indicated in the bottom half of Table 5, which he also attributes to AGR. When AGR is present, passivization is possible, since the AGR-governed subject is structurally distinct from the verb-governed object position. Furthermore, oblique (instrumental and dative) subjects subcategorized by the verb are only possible when AGR is not present, since AGR excludes government of the subject position by other possible governors. Finally, when AGR is present, its referential, pronominal features need to be identified with a referent through co-indexing with the subject position, and so null pronominals with arbitrary reference (pro_{arb}) are not possible. Thus, in this account, the patterns of register variable co-occurrence in Table 4 are necessary consequences of a single, binary-valued parameter.

In terms of communicative function, Literary Sinhala is the variety used for virtually all written Sinhala, not just literature (Gair 1986: 324). Exceptions are written communications in which the Literary endings would have too great a distancing effect, such as in personal letters to acquaintances, and in representations of speech, such as conversation in novels. In these contexts, and in ordinary conversation itself, some variety of Colloquial is used (Gair 1986: 324, 327–329). Formal speaking situations, in which public presentation or discussion is carried out in an impromptu or seemingly impromptu manner (such as sermons, university lectures, conferences and panel discussions on broadcast radio), use Formal Spoken Sinhala, which Gair considers a variety of Colloquial that happens to use Literary lexical items. Formal Spoken may also be used in personal letters (Gair 1986: 327–329). Thus, although Gair (1968, 1986) recognizes three registers of Sinhala, there is a strictly binary distinction in grammar, pivoting on the agreeing finite verbs of Literary. The two sub-varieties of Colloquial are claimed to be grammatically identical, and are distinguished only by lexical choice.

The different register variables of Literary and Colloquial Sinhala are thus claimed to be logically dependent upon one another, and to co-vary in a

[13] In equational sentences with third person subjects, the copula (or AGR) may be realized as the predicate marker *-yi/ya/yæ* (Gair 1992: 186).

[14] Gair's account requires that another governor be present that assigns accusative case to the subject when agreement is not present, otherwise the subjects of non-agreeing verbs would all have to be PRO. Gair does not specify what this governor is, though TENSE is not a viable alternative, since many non-finite verb forms lack tense but still may have overt (accusative case) subjects.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

categorical, bi-implicational relationship, so that the existence of any of the characteristics of Literary in Table 5 automatically implies the existence of all the others. This is an empirically falsifiable claim, should additional registers be found with mixed combinations of the Literary and Colloquial register variables of Table 5. In fact, a different, non-parametric account of the same phenomena is given in Paolillo (1991), in which the register variables above the hairline in Table 5 are claimed to vary more or less independently. As evidence of this, data from Sinhala literacy training texts are presented that have mixed combinations of the same register variables. It is reasonable to ask, therefore, which of these claims is correct – do the register variables of Literary and Colloquial registers vary categorically in natural texts (in other words, not merely in literacy training texts), as suggested by the parametric account, or not? If they do not vary categorically, what in fact is the distribution of the register variables, and how can that distribution be accounted for?

4. Register variation in a corpus of Sinhala texts

In studies of diglossic languages, the analysis of the functions of register variation is generally done informally, using either tables of communicative functions (Ferguson 1959), or choice trees (as in Britto 1986 for Tamil). Both types of account tend to view H and L as discrete, non-overlapping, whole entities, and individual register variables have no special status within them. The question to be addressed in this section is whether or not individual register variables of section 2 have the same distribution as one another, that is, whether or not a treatment in terms of unitary varieties is justified.

Register variation, like other types of sociolinguistic variation, is not readily amenable to introspection or informant elicitation, since, as noted by Labov (1966), and Hymes (1972), native speakers tend to represent the community's sociolinguistic norms in a distorted way. The only way to ascertain the distribution of the register variables described in the previous section is to undertake an empirical study of their occurrence in natural examples of actual language use. To this end, I conducted a survey of 29 Sinhala texts, comprising approximately 57,000 words, collected from a variety of circumstances. The texts were selected from a body of materials which I was able to obtain on research trips to Sri Lanka in 1986–87 and 1988, with the goal of identifying a complete range of register variation in Sinhala.[15] Thus, texts which represented the fullest available range of

[15] In part because of the scarcity of Sinhala materials outside of Sri Lanka, and in part because of the irreparable and unanticipated loss of some of the materials in transit, including most of the audio recordings, the corpus is somewhat modest, and it under-represents spoken forms of the language.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

communicative functions, and texts representing the fullest available range of combinations of register variables, were selected to be included in the corpus.

Within a single text, the register used can shift between any quoted speech or thought and the matrix text. A common example of this pattern in Sinhala occurs in literary novels, where conversation is represented in Colloquial, while only Literary is used for description and narration (De Silva 1967, 1974; Gair 1968, 1986; see also Blom & Gumperz 1972; Bell 1984). Thus, register variables are like other deictic forms in that their meaning varies according to who is the speaker, who may sometimes be the represented speaker of an embedded quotation. Where register variables differ compared to other types of deixis is that they attribute to the speaker a kind of communicative stance. Characters in a novel are represented as being 'authentic' (Britto 1986) through their use of colloquial speech, unless they happen to be engaged in formal speaking, or writing, in which case other register variables are used to represent the communicative stance appropriate to that kind of communication.[16] In other words, the communicative acts of literary characters are represented in a way that mirrors the register variation in the language as a whole.

In order to take this property of register variation into account, the texts of the corpus were segmented into units according to matrix and embedded discourse contexts. Quoted speech, thought, extended dialogues and text that was otherwise clearly the contribution of a different speaker were separated for purposes of analysis. The resulting corpus and its segmentation into text segments are described in the tables below. The text segments were examined for the occurrence of the register variables described above: agreeing finite verb forms, non-agreeing finite verb forms, accusative case, the accusative rule, locative case, non-finite copula verbs, and finite copula verbs. In the discussion below, groups of texts characterized by a similar pattern of register variables are presented together, followed by a general summary.

4.1 *Text segments with exclusively Colloquial variables*

There are four distinct general patterns of the distribution of the seven variables considered in this analysis: those which pattern exclusively like Colloquial for all variables, those which pattern exclusively like Literary, and two patterns of mixed Colloquial and Literary features (with several sub-patterns). Table 6 lists the text segments which exhibit the exclusively

[16] Actually, the situation is somewhat more complex; in a historical novel, drama or movie about classical-era subjects, high-status characters such as royalty and government officials tend to have more archaic Literary forms, including agreeing verb forms, in their speech. Lower status characters often show their uncultured nature (humorously) by using more Colloquial forms (De Silva 1967; Gair 1968, 1986). Similar patterns have been noted for other diglossic languages, such as Tamil (Britto 1986; Schiffman & Arokianathan 1986).

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

	verb		acc.			copula			
Seg.	agr	non	case	rule	loc.	non	fin.	words	Description
3.1	–	+	–	–	–	–	–	715	Radio advertisements dramatizing real-life situations.
4.2	–	+	–	–	–	–	–	645	Quoted conversation from a 2nd grade Buddhist primer.
4.4	–	+	–	–	–	–	–	4203	Narrative portions of a 2nd grade Buddhist primer.
10.1	–	+	–	–	–	–	–	387	Conversational exchanges in a popular children's book.
10.2	–	+	–	–	–	–	–	1191	Narrative portions of the above book.
11.1	–	+	–	–	–	–	–	243	Conversational exchanges in a first-grade story book.
11.3	–	+	–	–	–	–	–	671	Narrative portions of a first-grade story book.
12.1	–	+	–	–	–	–	–	823	Conversational exchanges in a second-grade story book.
17.1	–	+	–	–	–	–	–	284	Conversation in a story told orally by a monk (1).
17.2	–	+	–	–	–	–	–	1437	Narrative portions of a story told orally by a monk (1).
18.2	–	+	–	–	–	–	–	1889	Narrative portions of a story told orally by a monk (2).
23.1	–	+	–	–	–	–	–	3177	Conversational exchanges in a narrative poem.
11.2	–	+	–	–	–	?	–	32	Quoted inscription in a first-grade story book.
20	–	+	–	–	–	?	–	132	Personal letter by a young Sinhala woman (2).

Table 6
Text segments with exclusively Colloquial variables

Colloquial pattern. In this table, the first column contains the Segment number, for ease of reference. The next seven columns list the value of each variable for each segment: '+' indicates that the variable is present, '—' indicates that it is absent, and '?' indicates that the environment in which that variable should occur is not attested. For variables such as the accusative case and the accusative rule, '?' was only coded if the necessary environment did not occur in the text segment (for example, for the accusative case, a plural or indefinite animate or personal pronoun in object or prepositional object position). For other variables, such as the finite copula, it is not necessarily possible to tell whether the variable is absent or simply not attested on account of frequency. In these cases, a '?' was coded for the variable if the text segment was under 150 words in length, and '—' was coded otherwise. In Table 6, two text segments are coded this way: a very short (32 word) segment from a children's story book that is supposed to represent the inscription on a magical gemstone (segment 11.2), and a personal letter (segment 20). These are presented below the horizontal rule.

Although the group of texts in Table 6 appears to be somewhat heterogeneous at first, consisting of both written and spoken texts, closer inspection reveals that all but one of the written texts are composed with children as the primary audience. Since an important purpose of children's stories is literacy training, it makes sense that they would use predominantly Colloquial grammar, since Colloquial is the mother tongue of the children learning to read through them. The only other written text in Table 6 is the personal letter, a function which Gair identifies as sometimes using Colloquial. A preponderance of conversational exchanges is also represented in this group of text segments, including dramatic radio advertisements (segment 3.1), which contrast with non-dramatic advertisements, as will be seen below.

4.2 *Text segments with exclusively Literary variables*

Just as there are text segments whose grammar is exclusively Colloquial, there are others whose grammar is exclusively Literary; these are listed in Table 7. Included here are two types of patterns: one type in which there was solid evidence regarding all seven register variables (in the top portion of Table 7), and another in which evidence for one or more variables is lacking (in the bottom portion of Table 7). Since the hypothesis being considered claims that the exclusively Colloquial and exclusively Literary patterns should be the only two attested patterns, the Literary-like patterns with missing data will not be treated as disconfirming this hypothesis, even though they are consistent with other patterns found elsewhere in this analysis. This way, the parametric account will be disconfirmed only when the strongest possible evidence against it is available.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

	verb		acc.			copula			
Seg.	agr	non	case	rule	loc.	non	fin.	words	Description
1	+	−	+	+	+	+	+	3241	An article from a journal on cultural and linguistic studies.
2	+	−	+	+	+	+	+	270	An article summary abstract from the same journal as above.
5.1	+	−	+	+	+	+	+	2059	Quoted dialogue from a 6th grade Buddhist primer.
5.2	+	−	+	+	+	+	+	3093	Comprehension exercises from a 6th grade Buddhist primer.
5.3	+	−	+	+	+	+	+	9235	Narrative portions from a 6th grade Buddhist primer.
6.3	+	−	+	+	+	+	+	2082	The narrative portion of a book of reminiscences.
7	+	−	+	+	+	+	+	292	Coverleaf from a book on the history of Buddhism.
9.1	+	−	+	+	+	+	+	269	Introduction to 'The Prosperous Path of Divinity'.
9.3	+	−	+	+	+	+	+	1726	The Prosperous Path of Divinity' (pamphlet on morality).
14.1	+	−	+	+	+	+	+	232	Conversation from a story in a Buddhism text.
14.2	+	−	+	+	+	+	+	493	Narrative portions of a story in a Buddhism text.
22.1	+	−	+	+	+	+	+	871	Quoted speech of various participants in TV news.
22.2	+	−	+	+	+	+	+	1659	Descriptive and narrative portions of the TV news.
23.2	+	−	+	+	+	+	+	331	Introduction to a narrative poem.
23.3	+	−	+	+	+	+	+	3116	Narrative and descriptive verses from a narrative poem.
23.4	+	−	+	+	+	+	+	274	The publisher's preface to a narrative poem.
24	+	−	+	+	+	+	+	513	A pamphlet on socialism.
29	+	−	+	+	+	+	+	525	Word-by-word gloss of a Pāli religious text.
12.3	+	−	+	+	+	+	?	459	Author's preface to second-grade story book.
15	+	−	?	?	+	+	?	51	A monthly statement form from the People's Bank.
16	+	−	+	+	?	?	?	79	A subscription form for an academic journal.
26	+	−	+	+	?	?	?	14	Political graffiti on the Kelaniya University campus.
6.1	+	−	+	+	+	?	?	128	Author's foreword in a book of reminiscences.
25	+	−	+	+	+	?	?	50	Publicly posted signs and billboards.
9.2	+	−	?	?	+	+	+	199	Foreword to 'The Prosperous Path of Divinity'.
6.2	+	−	+	+	+	?	+	125	Publisher's foreword to a book of reminiscences.

Table 7
Text segments with exclusively Literary variables

All of the text segments exhibiting exclusively Literary register variables with no missing data are texts produced for public consumption which were scripted and edited in advance. Most of these materials were produced in written form, except for the television news program segments, which were scripted in advance of the broadcast and read on the air. The text segments with missing data vary in degree of spontaneity (for example, the graffiti vs. the bank statement) and in audience (for example, the preface to the second grade story book). It is possible that these factors influence the occurrence of the register variables, but the text segments also tend to be shorter, so it is at least equally likely that they do not provide sufficient opportunity to express the other variables.

4.3 *Text segments with mixed Colloquial and Literary variables*

There are three main groups of segments with mixed Colloquial and Literary register variables. One group of segments, given in Table 8, lacks agreeing verb forms. With the exception of two segments for which there is insufficient evidence on the locative, the five segments in this group share non-agreeing finite verb forms, accusative case, the accusative rule, and locative case. Two apparent sub-patterns can be discerned here with respect to the copula: one lacking any form of copula, the other lacking only the finite form of copula (segment 28 is indeterminate between these two patterns). In all of the text segments in Table 8, finite equational sentences are formed without *we*-, as in (7).

The communicative contexts of these five text segments may be characterized as either formal speaking (the linguistics lecture and the president's speech), sermonic (quoted monologue in the Buddhist primer, monthly Buddhist flyer), or both (the funerary sermon). This particular cluster of communicative functions is not entirely surprising, since Buddhist oration has historically had a profound impact on Sinhala oratory style, and on academic and political traditions as well. Nor should it surprise us that the well-recognized and statusful oratory patterns of Buddhist sermons would be seen to be suitable for some written purposes, even though most religious material uses exclusively Literary grammatical patterns.[17] Thus, it seems reasonable to characterize this group as exhibiting a 'Formal Spoken' grammar, although we must modify the description of Gair (1968, 1986) to note that this involves characteristics of Literary morphosyntax (register variables), not just lexis.

[17] The author's need to project a certain status may play a role here; the Buddhist pamphlet written by a lay person (segments 9.1, 9.2 and 9.3) uses exclusively Literary grammar, while the flyer (segment 8), written by a monk, uses mixed (Formal Spoken) grammar.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

	verb		acc.			copula			
Seg.	agr	non	case	rule	loc.	non	fin.	words	Description
21	−	+	+	+	+	−	−	536	Introductory portion of a lecture on historical linguistics.
4.1	−	+	+	+	?	+	−	234	Quoted monologue from a 2nd grade Buddhist primer.
8	−	+	+	+	+	+	−	1296	Monthly flyer on Buddhist doctrine.
27	−	+	+	+	+	+	−	958	Transcript of an oral funerary sermon.
28	−	+	+	+	?	?	−	50	Excerpt of a speech by President J. R. Jayawardhana.

Table 8
Text segments with mixed register variables, lacking agreement

A second group with mixed Colloquial and Literary register variables, but different from the group displayed in Table 8, has both agreeing and non-agreeing finite verb forms. This group is displayed in Table 9. Again, two principal patterns are observed: one lacking the copula verb (in which finite copula-less equational sentences like example (7) are found instead), and another in which the copula verb is found (in which there are no copula-less sentences).

In these segments, special communicative effects were generated by the use of the agreeing and non-agreeing verb forms. The switch from Literary to Colloquial register variables responsible for each of these special communicative effects functions much like codeswitching between two languages, and each of the special communicative effects noted here has been observed independently for any of several of language pairs (compare examples in Gumperz 1982; Myers-Scotton 1993). For example, segment 19, the second of two personal letters in the corpus, adopts a more distant authorial stance than the other letter (segment 20), by presupposing less common ground with the recipient. It begins with a formal salutation, with formal lexis and verb agreement in the first paragraph. Once the writer has satisfactorily introduced herself, in the middle of the second paragraph she shifts to more Colloquial grammar, using non-agreeing verbs and copula-less equational sentences. The switch from Literary to Colloquial metaphorically symbolizes the letter writer's changing social distance to the recipient, from a respectful to a more personal distance. A different communicative effect is observed in segments 13.1 and 13.2, both from the same story book. These segments exhibit a pattern of discourse-level highlighting with agreeing and non-agreeing verb forms, such that the narrative foreground uses non-agreeing verb forms, while the more backgrounded settings and descriptions use agreeing verb forms. And finally, the conversation from the monk's oral retelling of a traditional story (segment 18.1) exhibits a pattern of glossing – the dialogue is first quoted as if read (or memorized) from the religious texts that the story appears in, and is then paraphrased in Colloquial grammar. Not all of the dialogue is treated in this salient way, either; non-pivotal exchanges are simply given in Colloquial grammar, with non-agreeing verb forms. The only text segment in which there is no clear pattern in the use of Colloquial and Literary forms is segment 4.3, the comprehension exercises of the Buddhism text.

In all of these text segments, the accusative rule takes a particular form. When agreeing verbs are present, non-agreeing finite verbs behave like non-agreeing non-finite verbs for case assignment; that is, they take both objects AND SUBJECTS in the accusative case, as illustrated in examples (13) and (14), two comprehension exercises from a single lesson of the second grade Buddhism text (segment 4.3). Example (13) has the non-agreeing finite verb and an accusative feminine subject, while example (14) has an agreeing finite verb and a nominative subject. This pattern holds for all the text segments

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

	verb		acc.			copula			
Seg.	agr	non	case	rule	loc.	non	fin.	words	Description
13.1	+	+	+	+	+	−	−	1116	Conversational exchanges in a 6th grade story book.
18.1	+	+	?	?	?	−	−	562	Conversation in a narrative told orally by a monk (2).
4.3	+	+	+	+	+	+	+	769	Comprehension exercises from a 2nd grade Buddhist primer.
13.2	+	+	+	+	+	+	+	1668	Narrative portions of 6th grade story book.
19	+	+	+	?	?	+	?	183	Personal letter by a young Sinhala woman (1).

Table 9
Text segments with mixed agreeing and non-agreeing verbs

	verb		acc.			copula			
Seg.	agr	non	case	rule	loc.	non	fin.	words	Description
12.2	+	−	+	+	+	+	−	2182	Narrative portions of second-grade story book.
3.2	+	−	?	?	−	−	+	689	Radio advertisements with non-dramatic content.

Table 10
Remaining text segments with mixed register variables

that have both agreeing and non-agreeing finite verbs. In other words, accusative default case (the accusative rule) is 'projected' across sentences.

(13) wesak pooyadaa mahaa maayaa deewiya kohi
Vesak full.moon-day great Maya queen.ACC where.LOC
giyaa da?
go.PST Q
'Where did the great Queen Maya go on Vesak (month) full moon day?'

(14) boosat ladaruwa kawuru kawuru waḍaagattoo da?
Bodhisattva infant who.NOM who.NOM HOLD.PST.3PL Q
'Who all held the infant Bodhisattva?'

Two final text segments, the grammatical patterns of which are unique in the corpus, are listed in Table 10. Segment 12.2, the narrative portions of a second grade story book, is like other Literary text segments, except that it has copula-less equational sentences. The story segment 12.2 belongs to is an intermediate level literacy training text, and for this reason it deliberately employs a hybrid Literary and Colloquial grammar (compare the Colloquial grammar of the first grade story, segments 11.1 and 11.3, and the Literary grammar of the sixth grade Buddhism text, segments 14.1 and 14.2). Segment 3.2, the non-dramatic radio advertising, is somewhat more difficult to explain. It exhibits agreement and finite copula forms, but accusative, locative and the non-finite copula verb do not appear; in fact, the environments for accusative case seem not to occur at all. Yet there is ample opportunity for locative case to appear: expressions of location are heavily represented in this text segment, often naming places of business. This segment contrasts sharply with the dramatic radio advertisements, both in grammar (segment 3.1 is Colloquial) and in audience: the characters in a dramatic advertisement address one another, the non-dramatic advertisements address the radio audience.

4.4 *Summary of observed patterns of register variables*

As this section has shown, the register variables examined in the corpus exhibit a complex distribution. The full range of observed patterns is summarized in Table 11.

The observed register variation confirms the three main registers of Sinhala identified in Gair (1968, 1986), with several modifications. First, in addition to the three main registers – Literary, Colloquial and Formal Spoken – we must allow for a fourth set of varieties, labeled here as 'Hybridized Literary'. Second, Formal Spoken and Colloquial differ grammatically, and do not appear to be more closely related to one another than Formal Spoken and Literary, on the basis of the register variables observed here. Formal Spoken and Literary share the accusative and locative

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

verb		acc.			copula		
agr	non	case	rule	loc.	non	fin.	Register
+	−	+	+	+	+	+	Literary
+	−	?	?	−	−	+	
+	−	+	+	+	+	−	Hybridized Literary
+	+	+	+	+	+	+	
+	+	+	+	+	−	−	
−	+	+	+	+	+	−	Formal Spoken
−	+	+	+	+	−	−	
−	+	−	−	−	−	−	Colloquial

Table 11
Distribution of register variables in the corpus

cases, the accusative rule and the non-finite forms of the copula verb, whereas Formal Spoken and Colloquial share only the non-agreeing finite verb forms and copula-less equational sentences. Third, in contrast to the predictions of the AGR-based parametric account, agreement implies accusative case, but not vice versa. Likewise, the occurrence of the copula verb implies accusative case, but not vice versa, nor do agreement and the copula verb imply each other. Moreover, all text segments having the accusative observe the accusative default case rule. Thus, register variation in Sinhala can be seen to be organized around two main variables: agreement and accusative case. Minor register variation patterns are introduced by variation in equational and copula verb constructions, and by avoidance of some of the case variables. The AGR-based parametric account cannot account for these patterns of variation.

5. A FORMAL ACCOUNT OF REGISTER VARIATION

How can a formal account address these relationships? On the one hand, a formal account of register variation in Sinhala needs to address the observed regularities in the distribution of the register variables. On the other hand, it should address the relationship between the communicative functions of the different registers and the register variables they exhibit. The account proposed here will accomplish this by adopting the view that each register variable expresses a kind of social meaning, which is in part responsible for its distribution. This social meaning is indexical, in that it is anchored to participants in the speaking situation. It is also attitudinal, in that it represents the speaker's communication as being in a particular way, for

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

example, as interactive, edited, and the like. I use the term COMMUNICATIVE ATTITUDE to describe these kinds of social meaning.[18]

I propose here four different communicative attitudes that condition the distribution of the register variables. For agreeing finite verbs, I propose a communicative attitude 'edited', to reflect the fact that agreement occurs principally in edited discourse written or scripted for large-scale public consumption. Moreover, according to Gair (1986: 328), agreeing verb forms militate against a sense of spontaneity, and hence are unsuitable for occasions in which the audience is to be engaged. For non-agreeing verbs and the copula-less equational sentences, I propose an attitude 'interactive', since they are the only forms used in face-to-face interaction and wherever authentic face-to-face interaction is represented. Also, use of non-agreeing verb forms alongside agreeing forms carries a communicative effect like immediacy or interactivity. For the copula verb *we*-, I propose an attitude 'correct', since its use is not required in either Formal Spoken or Hybridized Literary, and it appears to add an increased degree of formality to both when it is used, suggesting a value of prescriptive correctness.[19] Finally, for the accusative and locative cases, I propose an attitude 'public', since they are found in Formal Spoken, Literary and hybridized Literary registers, whose primary distinguishing characteristic is that they are all public forms of communication. Every use of Sinhala must index itself with respect to (at least) the communicative attitudes of editedness, interactivity and publicness by choosing the register variables appropriate to one's social circumstances. Furthermore, marked choices generate implicatures about the speaker or the speaker's intent, such as humorous, socially inept, etc.

These proposed communicative attitudes can be incorporated into a formal analysis of register variation if we assume that they constitute part of each variable's lexical meaning. Thus the agreeing verb forms would have 'edited' as part of their lexical meaning, non-agreeing verb forms would have 'interactive', etc. The different register forms should co-occur freely as appropriate to their social context of use. In fact, there is one important exception, for which an explanation is required: the predicted combination of variables for edited/non-public communication does not occur (agreement without accusative case). Of the available text segments which are both edited and non-public (the two personal letters), one is entirely in Colloquial (no agreement, no accusative), while the other code-switches between Literary in the opening paragraph (agreement and accusative case) and Formal Spoken in the body (accusative case, no agreement). Thus, it appears

[18] There is a rough correspondence between this notion and the notion of 'communicative key' in Hymes (1972). Compare also the notion of 'social type', discussed in Hudson (1990: 64–5, 1997 chapter 7).

[19] Further analysis might reveal the finite and non-finite forms of *we*- to require different communicative attitudes, but for the purpose of the present discussion, it will be assumed that only their grammatical conditioning is distinct.

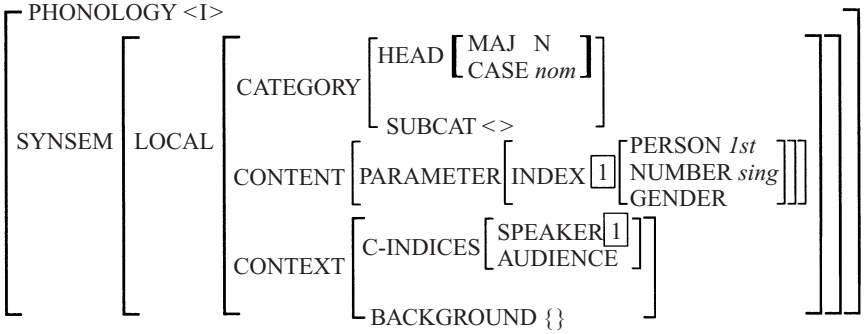
https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

that one cannot readily express the non-public, edited combination of communicative attitudes – another, similar combination (or combinations) must be chosen instead. I assume in the analysis below that the reason for this has to do with a language-particular grammatical dependency between agreement and accusative case. Acceptable alternatives to expressing the associated social meanings together are codeswitching (as in segment 19), in which editedness and non-publicness are expressed alternately by different variables, and avoidance of the contexts in which a variable would occur (as in segment 3.2).

5.1 *HPSG and unification*

We now have the prerequisites to propose a formal account of the register variables that addresses both their grammatical co-occurrence and their specialization for particular communicative functions. The account proposed in this section treats the communicative attitudes of register variables as part of their lexical entries, so that their grammatical distribution and social meaning are stated together in a coherent, unified analysis. A theoretical framework that allows this is the unification-based framework of Pollard and Sag (1987, 1994) known as Head-driven Phrase Structure Grammar or HPSG. HPSG is a formal theory of grammar (or more correctly a theory of language) whose principal mechanism for expressing relationships in linguistic expressions is the unification of attribute-value feature matrices. Attribute-value feature matrices are hierarchically structured collections of information that express the linguistic properties of a given morpheme, word, phrase, sentence or other component of language. For instance, the feature matrix given in Pollard & Sag (1994: 77) for the English pronoun *I* is given in example (15), showing the basic elements of the HPSG sign.

(15) *The lexical sign for the English pronoun* I



https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

Example (15) can be read as saying that *I* has two main components of information, its phonology (a list of phonemes indicated by the angle brackets), and its syntactic and semantic information, or SYNSEM feature (phrasal signs have an additional attribute, the DTRS or daughters attribute). The value of the SYNSEM feature is itself another feature matrix, with LOCAL and NON-LOCAL features; only the LOCAL feature is shown here. LOCAL in turn has the features CATEGORY, CONTENT and CONTEXT as its components. Features such as LOCAL which are referred to often are typically indicated by a shorthand 'path' name, such as SYNSEM|LOCAL, SYNSEM|LOCAL|CONTENT, etc.

The SYNSEM|LOCAL|CATEGORY of *I* is a feature structure indicating that *I* is a noun (its HEAD|MAJ feature has the value N) whose HEAD|CASE is nominative, with an empty SUBCATegorization list (in other words, it does not require any complements). The SYNSEM|LOCAL|CONTENT feature indicates it is a referential item, consisting of a PARAMETER whose INDEX has the referential features PERSON *1st* and NUMBER singular. The square box numbered 1 identifies this INDEX as being the same (token identical) feature structure that is pointed to by the CONTEXT|C-INDICES|SPEAKER feature , which is to say that this INDEX refers to the speaker in the context of utterance. The numbered box notation actually indicates a unification of parts of the feature structure bearing the same number (a 're-entrancy' in the terms of Shieber 1986). At times the specification for such a piece of information might come from more than one source; for example, the singular number of the subject in *she walks* comes from both the subject (*she*) and the verb (*walks*). Unifications in the representations of lexical items ensure that such information will be the same, no matter where it comes from. Alternatively, when a form such as the infinitive *walk* is compatible with multiple gender/number specifications (for example, *she might walk*; *I might walk*), the actual agreement features can be left unspecified in the lexical entry so that it unifies with the value of any appropriate subject. Ungrammaticality in a unification-based grammar is modeled as unification failure, which occurs when conflicting information from different sources is required to be unified but cannot be (as in **I walks*, with a third person singular verb and a first person singular subject).

Pollard & Sag recognize certain complex feature values, the most important of these being lists and sets. Lists, written between angle brackets ⟨ ⟩, are ordered feature values like the SUBCAT(egorization) list of a verb. The SUBCAT list lists the verb's arguments in the order of their relative 'obliqueness', so that the subject argument comes first, followed by the object, etc. Sets are used when unordered multiple values are required for a given feature. The standard curly bracket notation for sets is used to indicate set feature values. The set-valued feature that is most important for this study is the CONTEXT|BACKGROUND feature, which is used to specify the felicity conditions associated with a linguistic form.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

5.2 *Communicative attitudes and the* CONTEXT|BACKGROUND *feature*

Semantically, the communicative attitudes associated with register variables are felicity conditions; use of a register variable in an inappropriate circumstance (such as speaking in casual conversation using agreeing verb forms in Sinhala) results in an infelicity rather than an ungrammaticality or a semantic contradiction or falsehood. Furthermore, more than one communicative attitude may be relevant to any given speaking situation. The appropriate location for this type of information in HPSG, therefore, is in the BACKGROUND feature sets of the relevant lexical entries (Pollard & Sag 1994: 27, 91). Additionally, the BACKGROUND conditions governing the register variables are indexical and make reference to the speaker and audience of a speaking situation, much as the honorific agreement forms in Korean analyzed by Pollard & Sag (1994: 92ff.). These characteristics are made available through reference to the SYNSEM|LOCAL|CONTEXT|C-INDICES feature specification. Furthermore, the register variables make reference to the message itself, by representing it as more or less formal, public, or edited, etc. This can be accomplished through reference to the SYNSEM|LOCAL|CONTENT feature specification.

Pollard & Sag (1994: 332–337) discuss ways to handle the inheritance of BACKGROUND feature specifications by phrases, covering issues of both presupposition inheritance and perspectival (point-of-view) coherence, both of which are relevant to the analysis of register variation, as explained below. They do not, however, settle definitively upon a formal solution to the handling of CONTEXT for HPSG, noting that the issues raised thereby are the subject of active debate in formal semantics. We must therefore leave the precise formal statement of a CONTEXT feature principle to future research, while suggesting a provisional account to be used in the analysis below. The provisional account needs to address three main issues of register variation. First, different sentences of the same discourse use similar register variables. Second, direct quotations do not need to cohere in their choice of register variables with the matrix clause in which they are quoted. And third, special communicative effects can be generated through creative use of register variables, such as instances of humor signaled by the juxtapositions of register variables.

The first of these issues must be handled through the mechanism by which phrases inherit the BACKGROUND feature values of their daughters. Pollard & Sag (1994: 333) suggest a contextual consistency principle as a first-order approximation for handling BACKGROUND feature inheritance from phrases, under which the BACKGROUND value of a given phrase is the set union of the BACKGROUND values of the daughters. Pollard & Sag acknowledge this statement to be inadequate for many well-known problems of presupposition inheritance, but do not formalize an alternative. For the present purposes, I assume that phrases up to the sentence level generally inherit the

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

BACKGROUND feature values consisting of their daughters through set union; in the discussion below, I will refer to this informal principle as 'Normal BACKGROUND Inheritance'. Beyond the level of the sentence, in the discourse representation, set union of BACKGROUND feature sets may apply if they arise in temporally adjacent locations in the discourse, and their union does not lead to any contradictions; I will call this informal principle 'Discourse BACKGROUND Coherence'.[20]

Regarding the second issue, I assume that set union of the BACKGROUND features of phrases does not apply in cases where a set of contextual indices (CONTEXT|C-INDICES) distinct from those of the matrix clause is introduced. Such is the case in quoted speech with verbs of saying, where the embedded, quoted material need not have the same speaker and addressee as the matrix clause. In other words, when the speaker and addressee are distinct, I assume that the BACKGROUND feature sets may also be distinct (compare Sells 1987); in keeping with the above designations, I will call this 'Exceptional BACKGROUND Inheritance'.

Regarding the third and final issue, I assume that inferences of various sorts, especially implicatures, are calculated on the basis of the membership of the background feature set, which I term 'Discourse BACKGROUND Inference'. Although rules for Discourse BACKGROUND Inference are to be stated in formal terms, we should interpret them as specifying norms of communication (Hymes 1972) rather than aspects of the grammar per se. These same norms of communication are responsible for the fact that register variables with a value of, for example, *public* are found in communicative situations that are in fact public. Some of these norms would come into play in instances where adjacent sentences have contradictory BACKGROUND sets that cannot undergo union by Discourse Coherence. Implicatures generated by these norms include discourse highlighting functions (as in segment 13.2), and humor.[21] Thus, the analysis of BACKGROUND inheritance posits four principles: Normal Inheritance, for sharing BACKGROUND feature sets at the sentence level, Discourse Coherence, for the same at the discourse level, Exceptional Inheritance, for embedded communications, and Discourse Inference, for implicatures generated by BACKGROUND feature sets.

[20] The handling of BACKGROUND feature values at the sentence and discourse levels in HPSG appears to be a gap in the theory which is difficult to fill adequately. An anonymous reviewer notes that the formulation here could require computation of complex chains of inference which may turn out to be undecidable, and that a resource-limited version of consistency-checking (allowing some instances of inconsistent BACKGROUND specifications) might be both more psychologically realistic and more tractable.

[21] This account assumes a bisociation model of humor; see Attardo (1994) and Raskin (1985).

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

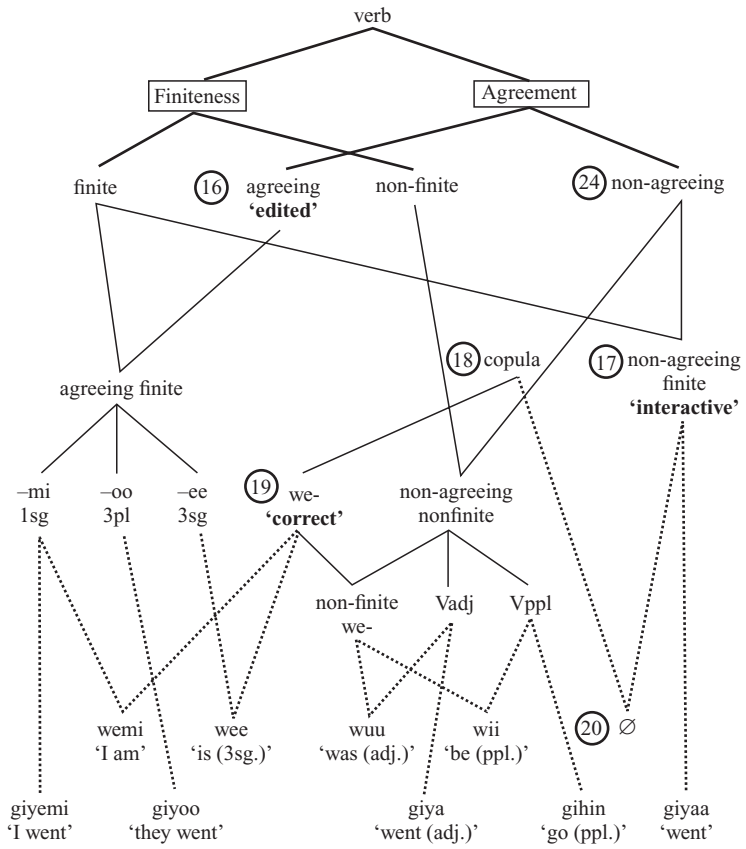


Figure 1
Lexical hierarchy for Sinhala verbs

5.3 *Agreeing and non-agreeing verb forms*

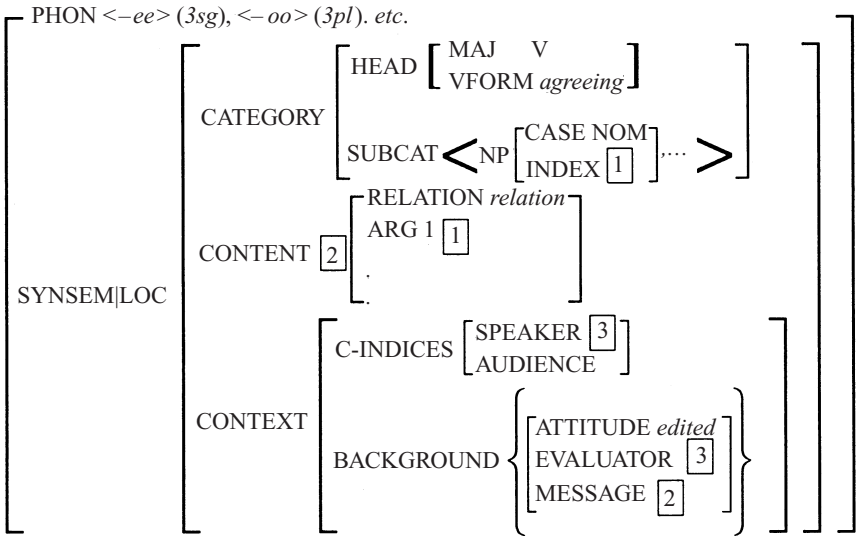
We can now stipulate a set of lexical forms for the Sinhala register variables described above, beginning with the agreeing and non-agreeing finite verb forms. These are best appreciated in the context of a multiple inheritance lexical hierarchy that describes Sinhala verb forms, given in Figure 1. In this hierarchy, nodes pertaining to tense and other inflectional categories are omitted for clarity. Types important to register variation have a communicative attitude indicated in bold; these are the types *agreeing*, *non-agreeing finite* and the *copula verb we-*. Types important to the HPSG analysis are marked with a circled number indicating the example number for the relevant feature matrix.

We begin with the agreeing verb forms, which are characterized by an 'edited' communicative attitude. The feature specification given in (16)

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

belongs to the *agreeing* type, which is unified with *finite*, and ultimately the feature specification of the verb stem. It specifies that the first argument on its SUBCAT list has nominative case, and that the INDEX feature of this argument is shared with the ARG1 feature of the relation in its own CONTENT feature, indicating that this argument is selected as subject.[22] The *agreeing* type also bears the communicative attitude *edited* in its BACKGROUND feature set. The EVALUATOR and MESSAGE features of this communicative attitude are unified with the SPEAKER contextual index and the CONTENT feature value, respectively.

(16) *Feature specification of* agreeing *verb type*

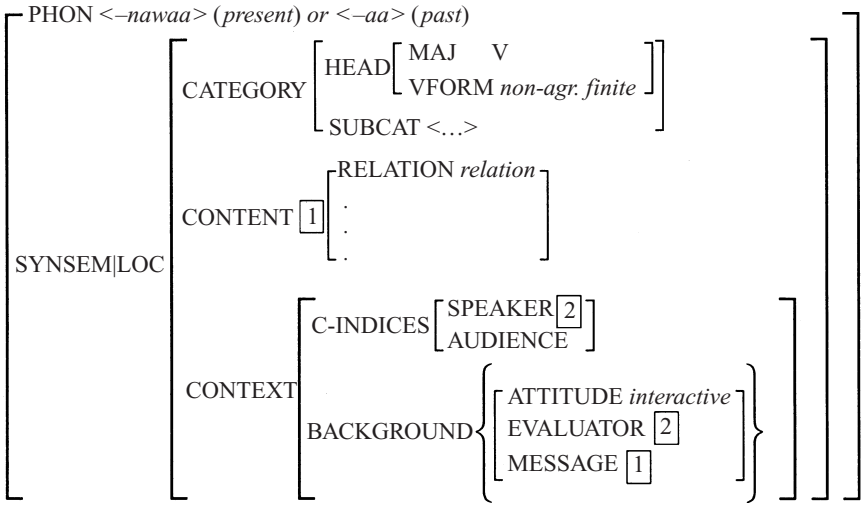


A similar feature specification with contrasting BACKGROUND feature values can be posited for the *non-agreeing finite* type. This feature specification is given in (17). Note that it is not necessary to state the subject case selection for the non-agreeing finite form, since a default case mechanism (described below) will provide this specification. The BACKGROUND feature set member *interactive* corresponds to the *edited* value posited for agreeing verb forms.

[22] For a more developed theory of argument selection by agreement morphology, see Pollard & Sag (1994), chapter 2. Since argument/subject selection is not the main interest here, an approximation will suffice.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

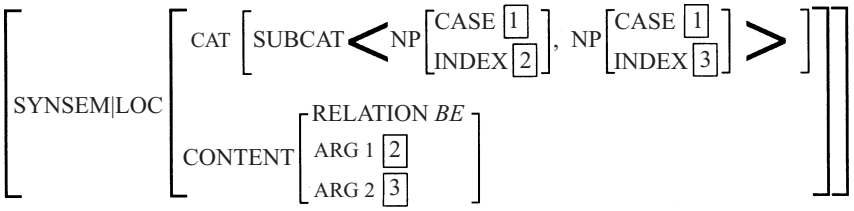
(17) *Feature specification of* non-agreeing finite *verb type*



5.4 *Copula verb and equational sentences*

The communicative attitudes of the finite and non-finite forms of the copula verb are relatively straightforward to represent, in the manner sketched for agreeing and non-agreeing finite verbs above. The grammar of the copula forms is more complex, since both the finite and non-finite copula verb forms have the unusual property that both arguments share the same case feature. All of the copular constructions share certain semantic and syntactic properties, which I assume are best represented by a type-raising predicate *BE* (Chierchia 1984). Analyzed this way, copular constructions have the feature specification indicated in (18).

(18) *Feature specification of* copula *type*

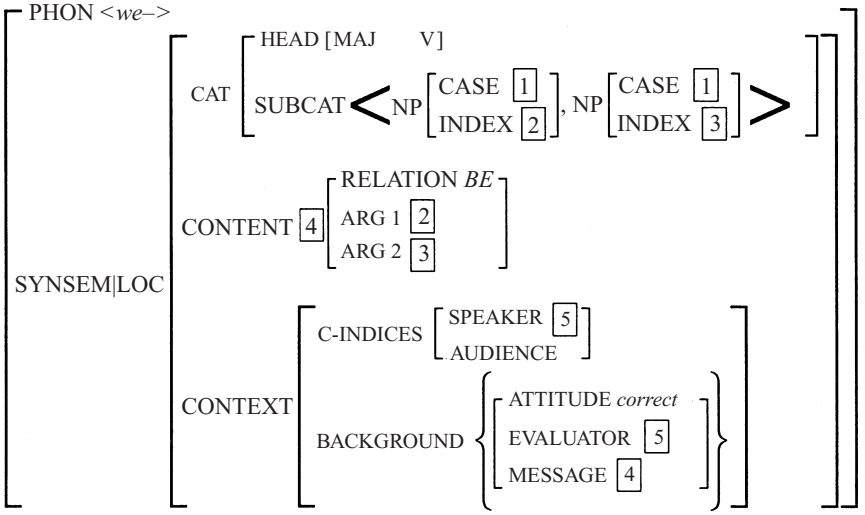


In (18), the case features of the two nominal arguments of the copula are shared, and the INDEX features of the two arguments are shared with the ARG1 and ARG2 of the CONTENT feature. These characteristics are common to all Sinhala copular constructions, whether a verb is present or not. The copula verb *we-* inherits this specification, and adds to it the communicative attitude

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

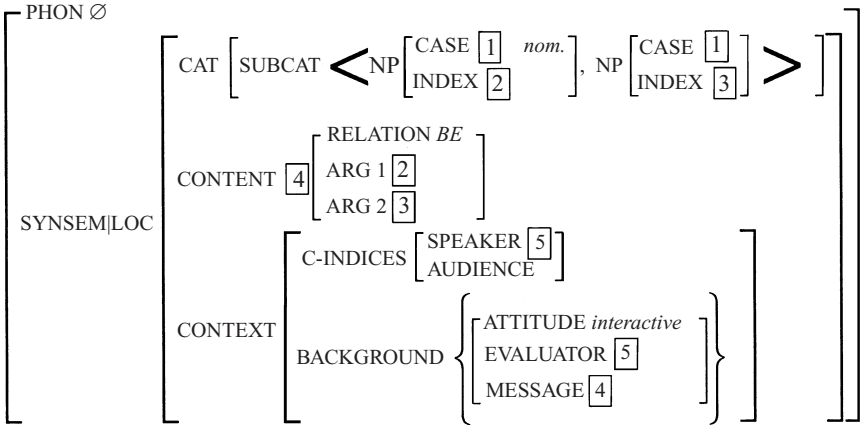
correct, as well as the CATEGORY ± HEAD and PHON feature specifications, as indicated in (19).

(19) *Feature specification of copula verb* we-



Copula-less equational sentences are a somewhat different matter, because there is no apparent lexical head. The way this is addressed here is to posit a 'zero' copula form, which has a null PHON specification.[23]

(20) *Feature specification of* 'zero' copula

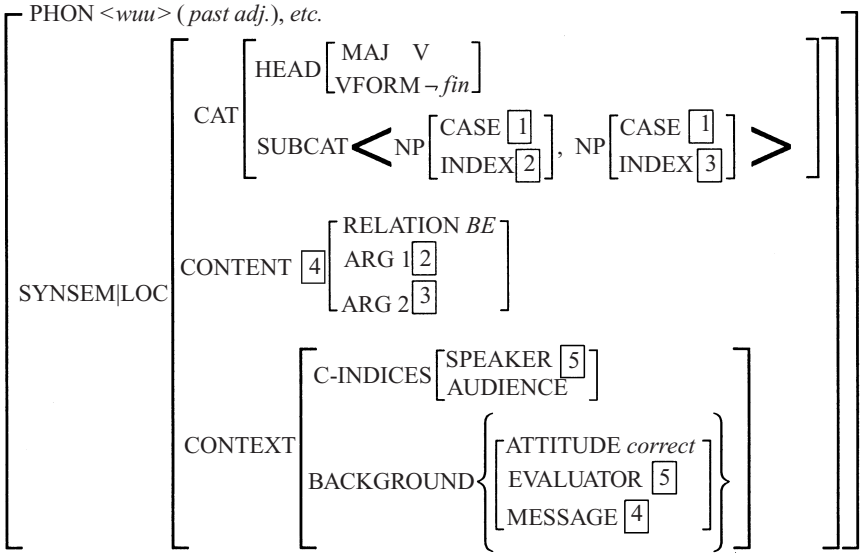


[23] An alternative to the zero copula analysis would be to posit a non-headed, phrasal, copular construction type, as would be required in Construction Grammar (Kay & Fillmore 1999). This approach substantially complicates the statement of the information common to both copulaless equational sentences and those with verbs, and so will not be pursued here.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

Non-finite forms of the copula verb simply inherit their characteristics from the copula verb *we-*, as indicated in (21).[24]

(21) *Feature specification of* non-finite copula *verb forms*



5.5 *Accusative, locative and the accusative rule*

Figure 2 illustrates the aspects of the lexical hierarchy for nouns relevant to register variation, with regard to the partitions of *Formality* (formal and non-formal) and *Case* (default, nominative, locative, genitive, dative, instrumental). Details of gender, number and definiteness have been omitted, so that the default case mechanisms can be more clearly represented.[25] Example nouns at the bottom of the diagram illustrate the surface manifestation of these forms.

The right hand side of Figure 2 consists of the partition relevant to case; each of the types under this partition assigns a value to the SYNSEM|LOCAL|CATEGORY|HEAD|CASE feature of the noun. Two types are somewhat exceptional: locative, which inherits from the type *formal* as well (to be explained presently), and *default*, which requires that the noun's

[24] An anonymous reviewer notes that the negation in the VFORM specification of (21) could be avoided by positing a type hierarchy for VFORM specifications.

[25] Also omitted are the details of inheritance for the surface genitive, dative and instrumental cases, all of which have both formal and non-formal variants (that is, they inherit from either the formal or non-formal type), with no difference in overt form.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

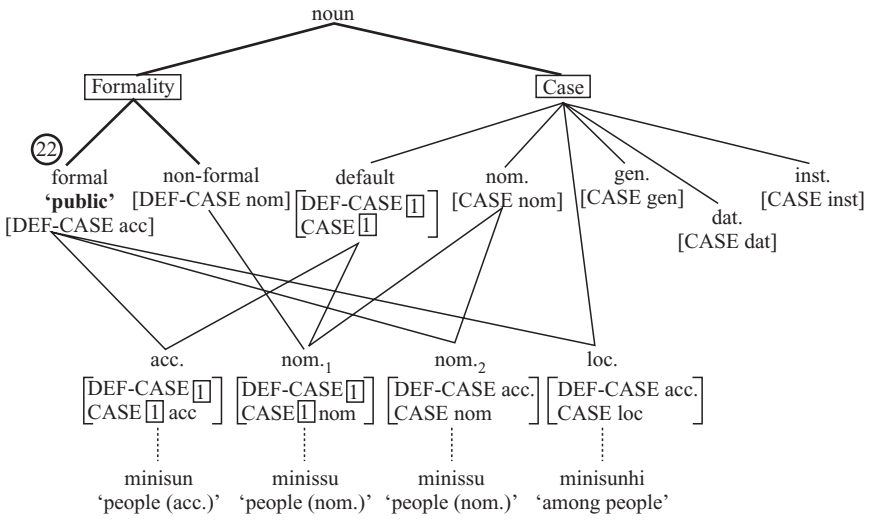


Figure 2
Lexical hierarchy for Sinhala nouns

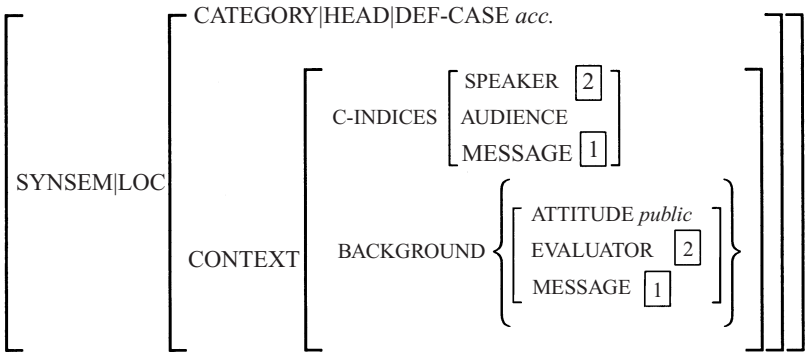
SYNSEM|LOCAL|CATEGORY|HEAD|CASE feature be shared with its SYNSEM|LOCAL|CATEGORY|HEAD|DEF(ault)-CASE feature; in other words, it requires that a noun appear in a default case. It is assumed that the *default* case type is specified in the subcategorization lists of postpositions and verbs, when no other case is specified. It has two alternative realizations, namely *accusative* and *nominative*$_1$. These realizations differ as the result of unifying *default* with one of the Formality types. Unification with *formal*, whose DEF-CASE specification is accusative, results in the type *accusative*. Unification with *non-formal*, whose DEF-CASE specification is nominative, results in the type *nominative*$_1$. Another type, *nominative*$_2$, is also permitted as the unification of *nominative* with *formal*. It is *nominative*$_2$ which appears as the subject of agreeing verb forms, since neither instance of *default* would satisfy the requirement of distinct DEF-CASE (= acc.) and CASE (= nom.) features.

It might seem at first that a feature specification such as DEF-CASE lends itself to a less elegant analysis than one that employs a version of default inheritance in the lexicon. However, such a solution would not suffice here, since there is not a single default case system, and in order for the analysis to work, the default case information needs to be shared both within and across sentences. Moreover, the mere existence of the accusative or locative case in a register is all that is needed to require that other default case nominals appear in the accusative. Default inheritance only helps when the system of default assignment does not vary across contexts, and when default information does not need to be shared syntactically. For these reasons, a default case feature must be posited.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

Of the different types identified in Figure 2, only one needs to bear a communicative attitude specification, namely the type *formal*. This type is more fully detailed in (22).

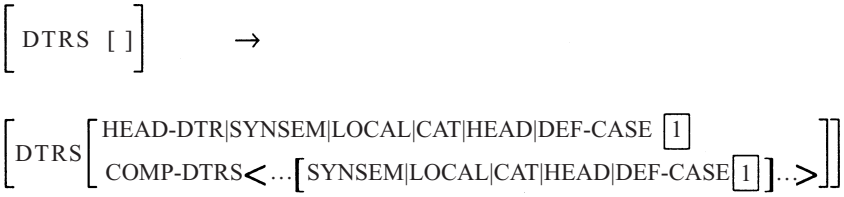
(22) *Feature specification for the type* formal



With these assumptions, the accusative rule is given the following analysis. The feature DEF-CASE is shared syntactically among the nominals within a sentence, through a *Default case principle*, specified in (23). This ensures that all nouns will share the same value for default case, and hence, that any that do appear in the default case will all appear in the same case (a mismatch of case among default case nouns will cause unification failure). Note that since DEF-CASE is a head feature (it is part of the SYNSEM|LOCAL|CAT|HEAD feature matrix), the Head Feature Principle (Pollard and Sag 1987: 58; 1994: 34) applies to it, so the DEF-CASE feature of a phrasal sign will also be shared with its head.

(23) *Default case principle*

The head of a phrasal node shares its default case feature with all its sister complement nodes.



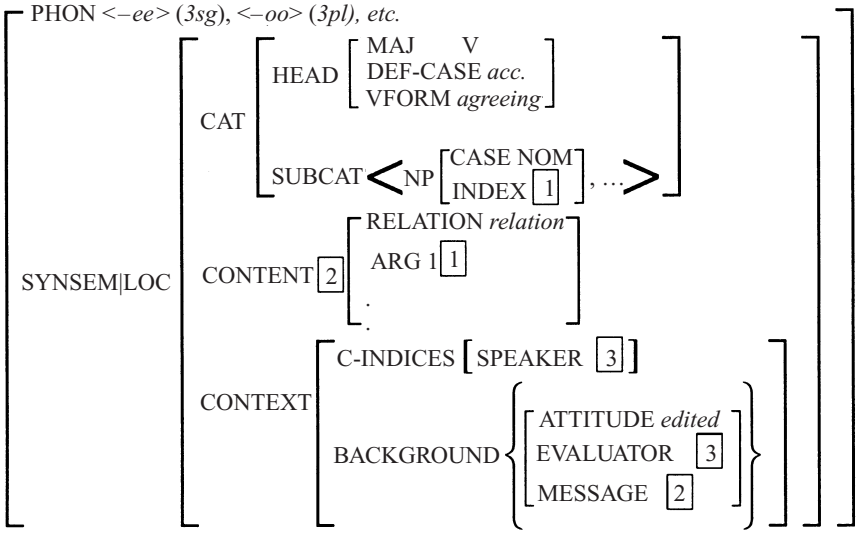
We now need only the additional stipulation that agreeing verbs have a DEF-CASE specification of accusative, to ensure that the default case for agreeing verb forms will not be nominative.[26] This stipulation is given in

[26] An anonymous reviewer suggests that it is counter-intuitive to require a default-case specification on verbs, when case is a feature of nouns. However, the problem that the analysis addresses is the strict agreement of all default case values within the sentence. This

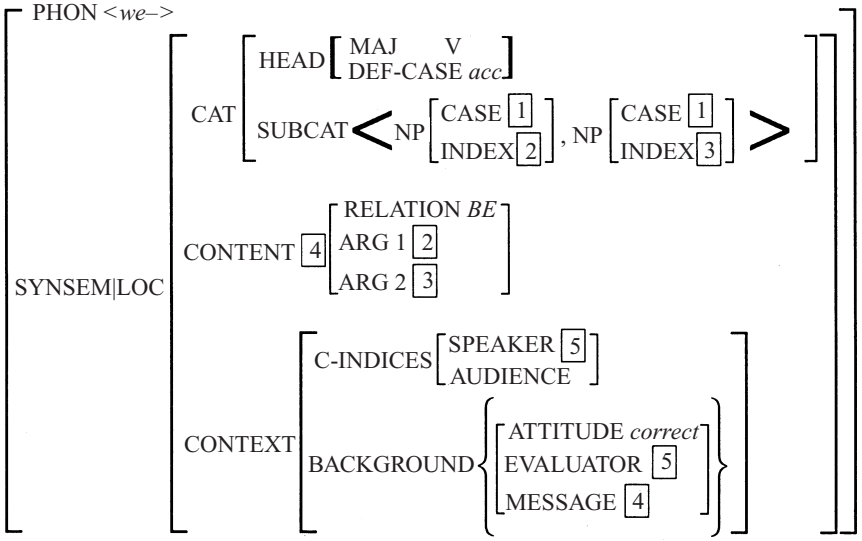
https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

(16′), which is a slightly modified version of (16). Likewise, co-occurrence of forms of the copula *we-* with accusative default case is indicated by a default case specification, as in (19′)

(16′) *Feature specification of* agreeing *verb type* (*revised*)



(19′) *Feature specification of copula verb* we- (*revised*)

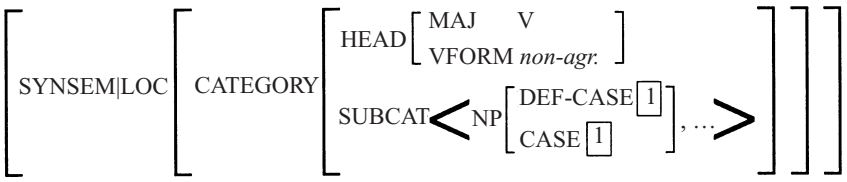


agreement involves the case-assigning properties of all the governors of nominals within the sentence (verbs and postpositions), at the very least. I would welcome any suggestions of a more elegant default-case mechanism that could handle the observed agreement facts.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

The default case specification of non-agreeing verbs can be left unspecified, as indicated in (24), so that they are compatible with a DEF-CASE specification of either nominative (as in Colloquial register) or accusative (as in Formal Spoken register). The CASE specification of the subject argument in the SUBCAT list must be defined as type *default*, so that we can account for the behavior of the subject of non-agreeing verbs as a default case position according to the accusative rule.

(24) *Feature specification of* non-agreeing *verb type*



The analysis of default case thus works in two modules: lexical and syntactic. First, agreeing verb forms are specified in the lexicon as having accusative as the default case, while non-agreeing verb forms do not specify a default case, and nominal types are cross-classified according to case and formality, which introduces DEF-CASE features. Second, the default case principle in (23) ensures that all daughters of a phrasal sign will share the DEF-CASE specification of the mother. If either $nominative_1$ or *accusative* appears in a position assigned *default*, or if a *locative* nominal appears, all of the DEF-CASE features of the entire sentence will have to share the value *accusative*, otherwise unification failure results.

The account also extends to copular sentences. The 'zero copula' specifies nominative as the CASE feature of its first argument, which is also shared by the CASE feature on its second argument; hence, arguments of zero-copula constructions will be nominative.[27] In agreeing *we*- copula constructions, the first argument is assigned $nominative_2$ since agreement requires the first argument to be nominative, and since the agreeing verb's DEF-CASE value is accusative; that value must be shared with the DEF-CASE of any noun in the sentence according to the Default case principle in (23). Since the CASE

[27] The analysis of the zero copula indicated in Figure 2 requires a lexical inheritance scheme in which a specified value must 'override' another higher up in the type hierarchy. On the one hand, the zero copula needs to inherit the 'interactive' communicative attitude from the non-agreeing finite form. On the other hand, it must override the *default* specification of the subject argument with $nominative_2$, since zero copula constructions only appear with nominative arguments, even in registers with accusative default case. Such a default inheritance scheme is not part of HPSG proper (for consequences of this, see Green & Morgan 1996). This can be corrected in Figure 2 by further developing the analysis of the lexical hierarchy verbs by including verb transitivity and other zero-verb subcategorization types, an area of rich variation in Sinhala grammar (see Gair 1998) not otherwise relevant to the issues presented here.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

feature of the first argument is shared with the CASE feature of the second argument, again both will be nominative. Finally, in the case of the non-agreeing, non-finite forms of the copula *we*-, the default-case feature is not set, so they are free to share DEF-CASE feature values that are either nominative or accusative. As a consequence, in registers where either agreement, accusative, or locative case appears, both arguments of a non-agreeing, non-finite form of *we*- will be accusative.

5.6 *Register redefined*

Registers can now be seen to be constituted by systems of forms that are coherent both in grammar and according to the norms of communication governing a given communicative situation. This means that one can specify a social context and predict the grammar of the expressions used in that context. This can be done for the four observed registers of Sinhala as follows.

Literary: For a written, public context, such as for a newspaper article, agreeing finite verb forms will be selected, since they express 'edited-ness'. Accusative case is selected because it expresses 'publicness', and because it is required as the default case for agreeing verbs. Normal BACKGROUND Inheritance combines the 'edited' and 'public' communicative attitude values through set union, into a set that is shared with other sentences of the discourse by Discourse BACKGROUND Coherence.

Formal Spoken: In a formal speaking situation, such as a sermon, non-agreeing finite verbs (which are 'interactive') are selected instead of agreeing forms. Selection of agreeing forms (with the 'edited' value) would lead to infelicitous inferences by Discourse BACKGROUND Inference. However, since the situation is public, forms expressing 'public' are required for the same reason, meaning that accusative or locative will be used, making accusative function as the default case, by virtue of the DEF-CASE value. The copula verb (with its value of 'correct') is possible, but not required. Speakers have latitude in using it depending on what they wish to convey with regard to 'correctness'.

Colloquial: In an informal speaking situation, such as ordinary conversation, 'interactive' forms are selected, but forms expressing 'edited' and 'public' are infelicitous by Discourse BACKGROUND Inference. Nominative must therefore be the default case, and copular constructions must use the zero copula, since the alternatives would require accusative as the default case, resulting in expressing an attitude of 'public-ness'.

Hybridized Literary: Different varieties of Hybridized Literary exist, which make different selections among the possible grammatical features of Sinhala. The variety used in the radio advertisements utilizes a subset of

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

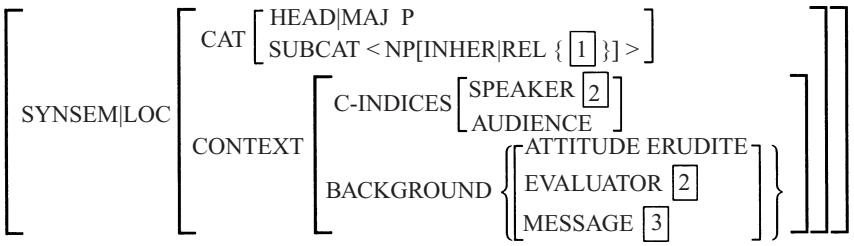
Literary features, deliberately avoiding use of accusative and locative case (by avoiding their environments) to avoid expressing 'public-ness'. By Discourse BACKGROUND Inference, this manages to express a smaller degree of social distance than would be possible with Literary, while retaining the expression of 'edited-ness'. A second variety chooses both 'interactive' and 'edited' forms, but through Exceptional BACKGROUND Inheritance, they result in the text being broken into spans with contrasting BACKGROUND feature sets. By Discourse BACKGROUND Inference, this results in implicatures of a 'highlighting' nature.

The same approach used here could also be extended to other register-sensitive phenomena, such as the grammar of pied-piping in English and Germanic mentioned in the introduction. Variation in pied-piping can be accounted for in HPSG by assuming that the grammar allows an element that licenses displacement of the preposition governing a *wh*-word or phrase. Such an element could be represented as an abstract type in the lexical hierarchy from which prepositions may inherit. There would be two such types, as in (25), which allows a relative pronoun (or a phrase containing a relative pronoun) as its complement, carrying the communicative attitude 'erudite', and (26), which does not allow relative pronouns in its complement (by virtue of the empty set specified on the INHER|REL feature).

(25) *Feature specification for prepositions disallowing pied-piping*



(26) *Feature specification for prepositions allowing pied-piping*



The communicative attitude value is then shared with the BACKGROUND feature of relative clause and the sentence as a whole, through Normal BACKGROUND Inheritance. The 'bookish' stylistic weight of relative clauses with pied-piping is thus a consequence of the communicative attitude in the BACKGROUND specification of the preposition that licenses pied-piping. Likewise, the highly colloquial German examples of Webelhuth (1992: 118)

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

could be explained by a different communicative attitude (such as 'hip') associated with a similar licensing element.

6. Conclusions

The analysis of register variation in terms of BACKGROUND feature specifications addresses the criticisms of the more usual treatments raised in section 1. By explicitly recognizing the relationship of register to grammatical structure, and by including that relationship in the scope of grammatical inquiry, it is possible to construct an integrated account of the grammar of register variables and the social parameters that constrain their use. This move has certain benefits, especially that of providing a more complete account of grammatical phenomena. It also has consequences for the enterprise of grammatical description.

One of these consequences is that what appear to be grammatical regularities of a very general nature, as in the AGR analysis of Gair (1992), turn out on closer inspection to require more complicated, language particular analyses, such as that represented by the feature structures in (16)–(22). This may appear on the surface to result in a loss of parsimony for the analysis, but this loss is illusory since the more general account cannot cover the full range of relevant facts. Thus, the approach offered here suggests something about what kinds of general grammatical explanations are tenable for the morphosyntactic facts of Sinhala. The distribution of accusative case and subject-verb agreement according to both grammar and register suggests that a bi-implicational account of these two register variables will not work, and that considerable refinement of the AGR account would be needed in order for it to address the observations made here.

A second consequence of the approach is that it allows the formalization of both the grammatical and social conditions relevant to register variation within a single, unified analysis. The use of HPSG obviates the need for separate theories or formalisms to express the two types of conditions. This particular advantage is not necessarily limited to HPSG analyses of register. Parallel approaches would be possible which adopt a hybrid theory having different components for the social and grammatical constraints. For instance, one might augment the Government and Binding approach employed by Gair (1992) with a Discourse Representation Theory analysis (Kamp 1990) for the felicity conditions on the register variables. However, the HPSG formalism has the advantage that it already encompasses both levels of description.

A further consequence of this approach concerns the object of grammatical analysis. While syntacticians and grammatical theorists often take themselves to be describing 'universal patterns in the grammar(s) of human language' or 'the grammar of a (particular) language', they typically privilege certain

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

forms of language over others; for example, structuralists generally consider the proper object of grammatical description to be only the spoken form of a language, since they regard the written language as derivative from and secondary to the spoken language, while generativists take grammaticality judgments to be the preferred object of study, since they allow properties of language to be tested that could be unaddressed or mis-represented in a linguistic corpus. However, both perspectives lead to a neglect of the kinds of regularities observed in the Sinhala data discussed here. Adopting the structuralist perspective would proscribe examination of the written language (and perhaps the formal spoken language as well), thereby obscuring and leaving unexplained the substantial morphosyntactic differences between Colloquial and Literary registers. Adopting the generative perspective wherein only grammaticality judgments are used leaves the analyst vulnerable to the notorious instability of native speaker judgments about sociolinguistic norms (Hymes 1972), which can only be reliably ascertained through observations of actual language use. Thus, accepting that grammatical analyses should address register variation means one also has to accept that grammatical analyses must be supported by non-introspective, empirical explorations of the language of different registers, beyond but inclusive of spoken language use.

A final consequence of this approach is that it provides a template for the analysis of other kinds of sociolinguistic variation. Other variable elements, whether they mark individual, social, ethnic or other dialect differences (Bell 1984, Labov 1966), genre or text type (Biber 1988), or differences between separate languages (Gumperz 1982, Myers-Scotton 1993), have indexical properties similar to those of the register variables studied here: they index attitudinal relations between speaker and audience of the communication, and they generate implicatures according to the norms and circumstances of their use. Thus, the account of register variation offered here can easily be modified to account for, for example, codeswitching between Swahili and English, by specifying the necessary BACKGROUND feature values and identifying the appropriate linguistic elements that bear them in that situation. Such an extension is consonant with the strongly pragmatic account suggested by Gumperz (1982: 91–99) for language-language codeswitching, and with the principle of 'audience design' invoked by Bell (1984) to explain other types of sociolinguistic variation.

Differences, as well as similarities, among different types of sociolinguistic variation can also be accounted for in this approach. A rough typology of variation can be offered by relating the different types to the general descriptive framework of Hymes (1972), known to many by the SPEAKING mnemonic. In Hymes' framework, communicative events must be described in terms of a broad range of factors, including the setting, the participants, the ends or purposes, the component speech acts, the communicative 'key', the instrumentalities or means of communication, the norms of com-

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

munication, and the genres or culturally recognized categories of communication involved in the event. Under this framework, inter-lingual code-switching and regional and social dialects are concerned principally with the social identities of the participants: who the speaker and hearer are, and their social relationship to one another. Variation by text type is concerned principally with the purposes and genres of communication. Register variation differs from other types of variation in that it principally concerns factors of the social setting, participants' roles within that setting, and the communicative key. Certainly there are relationships among all these types of variation – differences in setting can favor different media, purposes or genres, and the speaker-centeredness of register variation is similar to that of inter-lingual code switching and social and regional dialect variation. But the present approach also allows for different types of sociolinguistic variation to be represented by different types of BACKGROUND feature specifications which are subject to different norms of communication, so that analyses can be developed wherein, for example, factors relating to setting are more important in register variation, while factors relating to social identity are more important in dialect variation.

Thus a framework for the formal analysis of social variation in language emerges from this investigation, in which the requirements for describing and theorizing about register variation are not seen to be in conflict with the requirements of grammatical description and theory. Moreover, the theoretical tools required to address register variation (and by extension, other types of social variation) turn out to be theoretical tools required for other aspects of language, such as honorific agreement, presupposition and implicature. Formalizing register constraints on linguistic forms requires no theoretical machinery that is not already motivated for some other linguistic purpose. Finally, by attending to register variation, apparently simple grammatical patterns, such as those concerning the assignment of accusative case in Sinhala, are revealed to have a more complex, yet nonetheless highly structured, nature. A body of data organized according to register can reveal more clearly what the true limits of grammatical variation are. There is thus no reason not to address register variation in grammatical description and theory, and every reason to do so.

REFERENCES

Attardo, S. (1994). *Linguistic theories of humor*. The Hague: Mouton de Gruyter.
Bell, A. (1984). Language style as audience design. *Language in Society* **13**. 145–204.
Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
Biber, D. (1994). An analytical framework for register studies. In Biber & Finegan (eds.), 31–56.
Biber, D. & Finegan, E. (eds.) (1994). *Sociolinguistic perspectives on register*. Oxford: Oxford University Press.
Black, J. R. & Motapanyane, V. (eds.) (1996). *Microparametric syntax and dialect variation*. Amsterdam: Benjamins.
Blom, J.-P. & Gumperz, J. (1972). Social meaning in linguistic structure: code-switching in Norway. In Gumperz & Hymes (eds.), 407–434.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

Britto, F. (1986). *Diglossia: a study of the theory with application to Tamil*. Washington, DC: Georgetown University Press.
Chatterji, S. (1962). *A study of the relationships between written and spoken Bengali*. Ph.D. dissertation, Hartford Seminary Foundation.
Chatterji, S. (1986). Diglossia in Bengali. In Krishnamurti et al. (eds.), 294–303.
Chierchia, G. (1984). *Topics in the syntax and semantics of infinitives and gerunds*. Ph.D. dissertation, University of Massachusetts at Amherst.
Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
Culy, C. (1996). Null objects in English recipes. *Language variation and change* **8**. 91–124.
De Silva, M. W. S. (1967). Effects of purism on the evolution of written language. *Linguistics* **36**. 5–17.
De Silva, M. W. S. (1974). Convergence in diglossia: the Sinhalese situation. *International Journal of Dravidian Linguistics* **3**. 60–91.
De Silva, M. W. S. (1976). *Diglossia and literacy*. Mysore: Central Institute of Indian Languages.
Fasold, R. (1984). *Sociolinguistics of society*. Oxford: Blackwell.
Ferguson, C. A. (1959). Diglossia. *Word* **51**. 325–340. [Reprinted in Ferguson, C. A. (1996). 25–39.]
Ferguson, C. A. (1971). Absence of copula and the notion of simplicity: a study of normal speech, baby talk foreigner talk and pidgins. In Hymes, D. (ed.), *Pidginization and creolization of languages*. Cambridge: Cambridge University Press. [Reprinted in Ferguson (1996). 115–123.
Ferguson, C. A. (1994). Dialect, register and genre: working assumptions about conventionalization. In Biber & Finegan (eds.), 15–30.
Ferguson, C. A. (1996). *Sociolinguistic perspectives: papers on language and society 1959–1994* (T. Huebner, ed.). Oxford: Oxford University Press.
Fishman, J. (1967). Bilingualism with and without diglossia, diglossia with and without bilingualism. *Journal of Social Issues* **23.2**. 29–38.
Gair, J. W. (1968). Sinhalese diglossia. *Anthropological Linguistics* **10**. 1–15
Gair, J. W. (1986). Sinhala diglossia revisited, or diglossia dies hard. In Krishnamurti et al. (eds.), 322–336.
Gair, J. W. (1992). AGR, INFL, Case and Sinhala diglossia or: can linguistic theory find a home in variety? In Kachru, B., Dimock, E. C. & Krishnamurti, Bh. (eds.), *Dimensions of South Asia as a sociolinguistic area: papers in memory of Gerald B. Kelley*. New Delhi: Oxford & IBH Publishing Co. 179–198.
Gair, J. W. (1998). *Studies in South Asian linguistics* (B. Lust, ed.). Oxford: Oxford University Press.
Gair, J. W. & Karunatilaka, W. S. (1974). *Literary Sinhala*. Ithaca NY: Cornell University South Asia Program.
Gair, J. W. & Paolillo, J. C. (1988). Sinhala non-verbal sentences and argument structure. *Cornell Working Papers in Linguistics* **8**. 39–78. [Reprinted in Gair, J. W. (1998). 87–107.]
Gair, J. W. & Paolillo, J. C. (1997). *Sinhala*. Languages of the World/Materials, No. 34. München: Lincom Europa.
Green, G. M. & Morgan, J. L. (1996). Auxiliary inversions and the notion of 'default specification'. *Journal of Lingusitics* **32**. 43–56.
Gumperz, J. & Hymes, D. (eds.) (1972). *Directions in sociolinguistics*. New York: Holt, Reinhart & Winston.
Gumperz, J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.
Haegeman, L. (1987). Register variation in English: some theoretical observations. *Journal of English Linguistics* **20**. 230–248.
Henry, A. (1995). *Belfast English and Standard English: dialect variation and parameter setting*. Oxford: Oxford University Press.
Henry, A. (1996). Imperative inversion in Belfast English. In Black & Motapanyane (eds.), 79–93.
Hudson, A. (ed.) (1991). *Studies in diglossia*. Special issue of *Southwest Journal of Linguistics* **10.1**.
Hudson, A. (1992). Diglossia: a bibliographic review. *Language in Society* **21**. 611–674.
Hudson, R. (1990). *English word grammar*. Oxford: Blackwell.

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press

Hudson, R. (1997). *Sociolinguistics* (2nd edn.). Cambridge: Cambridge University Press.
Hymes, D. (1972). Models of the interaction of language and social life. In Gumperz & Hymes (eds.), 35–71.
Kay, P. & Fillmore, C. (1999). Grammatical constructions and linguistic generalizations: The *What's X doing Y?* construction. *Language* **75. 1–33.**
Kamp, H. (1990). *From discourse to logic*, vol. 1. Boston: Kluwer.
King, C. R. (1994). *One language, two scripts*. Bombay: Oxford University Press.
Krishnamurti, Bh. (1977). Language planning and development: the case of Telugu *Contributions to Asian Studies* **11**. 37–56.
Krishnamurti, Bh., Masica, C. P. & Sinha, A. K. (eds.) (1986). *South Asian languages: structure, convergence, and diglossia*. Delhi: Motilal Banarsidass.
Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
Longacre, R. (1996). *The grammar of discourse*. New York: Plenum.
Milroy, L. & Muysken, P. (eds.) (1995). *One speaker, two languages*. Cambridge: Cambridge University Press.
Myers-Scotton, C. (1993). *Duelling languages*. Oxford: Oxford University Press.
Paolillo, J. C. (1991). Sinhala diglossia and the theory of Government and Binding. In Hudson, A. (ed.), 41–59.
Paolillo, J. C. (1992). *Functional articulation in diglossia: a case study of grammatical and social correspondences in Sinhala*. Ph.D. dissertation, Stanford University.
Paolillo, J. C. (1997). Sinhala diglossic variation: continuous or discrete? *Language in Society* **26**. 269–296.
Pollard, C. & Sag, I. A. (1987). *An information-based theory of syntax and semantics*. Stanford: Center for the Study of Language and Information.
Pollard, C. & Sag, I. A. (1994). *Head-driven Phrase Structure Grammar*. Stanford: Center for the Study of Language and Information.
Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
Raskin, V. (1985). *Semantic mechanisms of humor*. Dordrecht: Reidel.
Ross, J. R. (1967). *Constraints on variables in syntax*. Doctoral dissertation, MIT.
Schiffman, H. & Arokianathan, S. (1986). Diglossic variation in Tamil film and fiction. In Krishnamurti, Bh., Masica, C. P. & Sinha, A. K. (eds.), 371–381.
Sells, P. (1987). Aspects of logophoricity. *Linguistic Inquiry* **18**. 445–479.
Shieber, S. M. (1986). *An introduction to Unification-based approaches to grammar* (CSLI Lecture Notes No. 4). Stanford, CA: CSLI Publications.
Trudgill, P. (1974). *Sociolinguistics: an introduction*. New York: Penguin.
Wardhaugh, R. (1992). *An introduction to sociolinguistics* (2nd ed). Oxford: Basil Blackwell.
Webelhuth, G. (1992). *Principles and parameters of syntactic saturation*. Oxford: Oxford University Press.

Author's address: *Program in Linguistics,*
Box 19559,
University of Texas at Arlington,
Arlington TX, 76019–0559,
U.S.A.
E-mail: john@ling.uta.edu

https://doi.org/10.1017/S0022226700008148 Published online by Cambridge University Press