SSLA, **27**, 535–566. Printed in the United States of America.
DOI: 10.1017/S0272263105050242

LANGUAGE TRANSFER AND DISCOURSE UNIVERSALS IN INDIAN ENGLISH ARTICLE USE

Devyani Sharma
King's College London

Stable nonnative varieties of English acquired and used in the absence of native English input can diverge systematically from native varieties over time (Cheshire, 1991; Kachru, 1983; Platt, Weber, & Ho, 1984). Focusing on Indian English article use, this study asks the following question: If divergence is indeed occurring, do new features derive primarily from first language (L1) transfer or from universal principles? Natural conversational speech is assessed in relation to four hypotheses relating to L1 transfer and language universals, and a multivariate regression analysis evaluates the relative strength of each factor. The new article system is not found to be identical to the L1 article system. Although L1 transfer appears to be operative when an overt form (the specific indefinite article) exists in the L1, when a gap occurs in the L1 (no definite article), speakers do not completely omit the definite article in their second language English. Using Prince's (1981) taxonomy of assumed familiarity, it is shown that the absence of a L1 model for definite articles permits the intervention of universally available discourse knowledge, such that speakers apply an economical, disambiguating principle to the use of overt articles, reserving them mainly for new (less given or inferable) information and omitting them in more redundant contexts.

Second languages (L2s) spoken in stable bilingual settings—for instance, English as spoken in India or Singapore—can become indigenized over time due to the historical removal of the native variety and broad societal trans-

I am indebted to John Rickford, Elizabeth Traugott, Arnold Zwicky, Ishtla Singh, and the *SSLA* editors and anonymous reviewers for much helpful feedback. I also received valuable comments from audiences at the LSA, Atlanta (January 2003) and at UC Davis and UC San Diego (February 2003).

Address correspondence to: Devyani Sharma, Department of English Language and Literature, King's College London, Strand, London, WC2R 2LS, United Kingdom; e-mail: devyani.sharma@kcl.ac.uk.

mission of L2 traits through institutional as well as informal channels. The forces that influence the direction of change in such cases are similar to forces active in classroom and individual SLA, but their ultimate impact might be much more lasting due to the increasing function of the L2 variety as a dialect in its own right (Kachru, 1983; Lowenberg, 1986; Platt, Weber, & Ho, 1984; Sahgal, 1991; Sridhar, 1985). As Kandiah (1991), Bamgbose (1998), and others have noted, if stable nonnative varieties of English indeed constitute a special case of SLA, then the distinctiveness of their grammatical systems must be shown through systematic analysis. The two primary goals of this study are to seek evidence of systematic nonstandardness in one such variety and to try to explain any such divergence. An implicational scaling of 12 Indian English speakers indicates higher rates of nonstandardness in article use as compared to other features, and the bulk of this study aims to identify the source of this new usage.

Explanations for SLA patterns in L2 English article use have included first language (L1) transfer (Jarvis, 2002; Platt et al., 1984), universal distinctions (Bickerton, 1981; Ionin, Ko, & Wexler, 2003), topic continuity (Chaudron & Parker, 1990; Huebner, 1983; Jarvis), and conversational genre (Tarone & Parrish, 1988). Although some of these factors have been suggested for New Englishes, they have not always been subjected to close scrutiny. The present study evaluates the relative importance of language transfer and linguistic universals in the development of new principles for article use. Within the domain of universals, a particular interest of this study lies in the reliance by L2 speakers on universally available knowledge about the discourse status of entities in the absence of a clear native target model. The findings strongly suggest that although language transfer plays a role—particularly when the L1 has an overt form designated for a particular function (as opposed to lacking a form that exists in the L2)—speakers appear to exploit universally available discourse knowledge to a significant extent as well as to create a new functional basis for the use of articles.

In the next section, four competing hypotheses that might account for new patterns of article use are proposed. The first three hypotheses have been examined in earlier studies, and the fourth introduces a new potential factor. The subsequent section describes aspects of the methodology, including the use of sociolinguistic interviews, a brief profile of the participants in the form of an implicational scaling (to demonstrate the greater degree of divergence in article use), and the methodology used for coding each article token. Each of the four hypotheses about article use is then evaluated and the relative importance of all factors is compared using multivariate regression analysis.

HYPOTHESES FOR INDIAN ENGLISH ARTICLE USE

Two core dimensions of crosslinguistic variation in article systems are specificity and definiteness. Specificity signals the existence of a unique real-world referent for a noun phrase (NP), or the "speaker's ability to identify the refer-

ent" (Fodor & Sag, 1982, p. 356). A nonspecific NP could refer to any token of its type in the real world (*I'm looking for a book. Any book will do)*, whereas a specific NP has a unique referent in the real world (*I'm looking for a book. I think I left it here yesterday*). Definiteness, by contrast, is primarily rooted in discourse. Hawkins (1978) described the definite article as an instruction for the hearer to locate the referent of that NP within a pragmatically defined set of objects that are part of the shared speaker-hearer knowledge. This can be referred to as the givenness of information that the speaker can treat as "recoverable either anaphorically or situationally" (Halliday, 1967, p. 211). A definite NP has already been referred to by a speaker and can be assumed to be known (*I asked a boy for directions. The boy told me*), whereas an indefinite NP has not been referred to by a speaker and cannot be assumed to be known (*I asked a boy for directions*).

Second language article use might be influenced by transfer of a L1 system of marking definiteness and specificity or by universal preferences in the linguistic marking of these categories. The first two hypotheses that will be presented relate to language transfer, and the latter two present potential universal factors.

Hypothesis 1: Transfer of L1 Positional Marking of Discourse Status

The Indo-Aryan L1s of the speakers examined in this study lack a definite article and mark definiteness via word order (McGregor, 1995), a system exemplified in (1).

(1) a. *kitaab* mez par paṛi hai
book table on lying is
"*The book* is lying on the table"
b. mez par *kitaab* paṛi hai
table on book lying is
"*A book* is lying on the table"

The definiteness interpretation of the NP *kitaab* in (1) varies according to its position in the clause. Discourse configurationality of this type frequently places information familiar to the hearer (thematic information) clause-initially and new (rhematic) information clause-finally. This crosslinguistically observed pattern has been variably described as topic/focus, topic/comment, and theme/rheme (Birner & Ward, 1998; Dik, 1978; Lambrecht, 1994; Vallduví, 1992). Thus, one type of L1 transfer in Indian English might be a reliance on clause position rather than articles to mark discourse status. In a study of English acquisition by speakers of Czech and Slovak—both discourse configurational languages—Young (1996) predicted that discourse items in unexpected (noncanonical) clausal positions would be overtly marked, and redundant marking would be avoided elsewhere. Given the similar discourse configurationality

of Hindi, I adopt Young's (p. 156) hypotheses as my hypothesis 1, presented in (2).

(2) Hypothesis 1 (transfer of L1 positional marking)
a. Thematic NPs in clause-initial position will favor zero articles. (unmarked)
b. Rhematic NPs in clause-final position will favor zero articles. (unmarked)
c. Thematic NPs in clause-final position will favor definite articles. (marked)
d. Rhematic NPs in clause-initial position will favor indefinite articles. (marked)

This predicts that speakers will use articles only when a NP is not in its pragmatically determined clause position (i.e., in situations (c) and (d)). Old information in the expected clause-initial theme position will favor zero articles, as will new information in the clause-final rheme position.

Hypothesis 2: Transfer of L1 Article System

Direct transfer from a distinct L1 article system is perhaps the most frequently cited explanation for divergence in article use in nonnative varieties of English. As a result of extensive areal convergence, the article systems of the five L1s in the present data are fairly similar: None of the languages has a definite article, definiteness can be marked by word order or by case-marking, and all of the languages can use the numeral *one* with specific indefinite meaning (Bhatia, 1993; Cardona, 1965; McGregor, 1995; Schiffman, 1999; Sridhar, 1990). In other words, the article systems of these languages contrast specific/nonspecific but not definite/indefinite. As the simplified diagrams in Figures 1 and 2 show, this system is the inverse of the article system of English, which prioritizes the definite/indefinite distinction over specific/nonspecific.

Platt et al. (1984, pp. 53–59) suggested that transfer of L1 specificity-marking is the primary factor in divergent article systems in New Englishes. These authors cited two Indian English examples, repeated in (3), that reflect direct adaptation of L2 forms—*one* and Ø—to L1 functions.

Definite	Indefinite	
the	a	Specific
the, a *(including generic)*	a	Nonspecific

Figure 1. English article system (for singular count nouns).

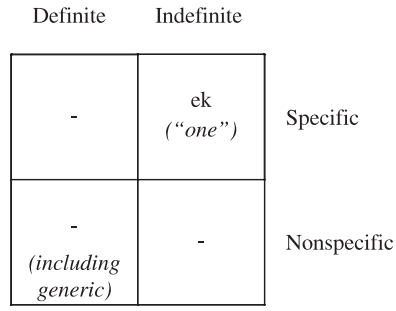

Figure 2. Hindi article system (for singular count nouns).

(3) a. I'm staying in *one* house with three other students. (specific)
b. I want to spend some time in Ø village, definitely if I get Ø chance. (nonspecific)

Two distinct predictions underlie transfer from a Hindi-like system (Figure 2) to English (Figure 1). First, the existence of a specific indefinite article in Hindi should lead speakers to restrict their English indefinite article use to specific NPs only. Second, speakers might avoid articles with definite and generic NPs in their English, as Hindi has no such marker. These predictions are presented in (4).

(4) Hypothesis 2 (transfer of L1 article forms and functions)
a. [+ specific, − definite] NPs will be marked overtly with an article (possibly using the form *one* rather than *a*).
[− specific, − definite] NPs will have no article.
b. [+ specific, + definite] NPs will have no article.
[− specific, + definite] NPs will have no article.

Hypothesis 3: Universal Tripartite Article System

A mismatch between the L1 and L2 grammars might permit the intervention of universally unmarked or more natural grammatical systems. Bickerton (1981) offered an explicitly formulated claim regarding universal distinctions in article use. He suggested that the semantic distinction of [± specific reference] and the discourse distinction of [± hearer-known] are universal and expressed by various means crosslinguistically. He argued that whereas the main parameter of the article system of standard English is definite/indefinite, Creole systems generate an additional specificity distinction to achieve a universally preferred tripartite system: (a) [+ specific reference, + hearer-known], such as referential definites, that require a definite article; (b) [+ specific reference, − hearer-known], such as referential indefinites, that require an indefinite article; and (c) [− specific reference, − hearer-known] or [− specific

reference, + hearer-known], such as nonspecifics and generics, that require zero marking.

Bickerton's (1981) model has been adopted and evaluated in several studies of article use in SLA and in new varieties of English. Chaudron and Parker (1990) found that Japanese speakers redistributed English forms to maintain the definiteness and topic continuity distinctions of [+ specific reference, + hearer-known, + topic], [+ specific reference, + hearer-known, − topic], and [+ specific reference, − hearer-known]. In his examination of a single individual's acquisition of the English article system, Huebner (1983) also found some evidence of Bickerton's categories; however, they were reflected in various developmental stages the learner moved through, rather than as a static, unified system. Mesthrie (1992, p. 205) suggested that sporadic article use in South African Indian English is also reminiscent of a prototypical Creole system. However, given that Bickerton would predict a distinct marker for each of the three types, the three examples Mesthrie furnished—which are shown in (5)—lend equal, if not more, support to a direct transfer of forms from the L1 system.

(5) a. I was feeling thirsty, so I bought *one* soda water. (+ specific reference, − hearer-known)
b. *Ø* Food is lovely. (+ specific reference, + hearer-known)
c. Because if they give us *Ø* chance. . . (− specific reference, − hearer-known)

The universalist or typological prediction for article use that uses Bickerton's (1981) distinctions of [± specific reference] and [± hearer-known] as a base can be summarized as in (6).

(6) Hypothesis 3 (universal prototypical tripartite division)
a. [+ specific reference, + hearer-known]: NPs will have a definite article.
b. [+ specific reference, − hearer-known]: NPs will have an indefinite article.
c. [− specific reference, − hearer-known] and [− specific reference, + hearer-known]: NPs will have no article.

Notice that of these predictions, (6b–6c) hold true for the article systems of Indian L1 languages, and hypothesis 3 is therefore identical in those respects to hypothesis 2; as we will see, this renders the distinction between transfer and universal influences potentially problematic.

Hypothesis 4: Universal Discourse Knowledge

A second type of universal influence on L2 article use could come from discourse pragmatics rather than semantic contrasts. The status of a NP as [+ hearer-known] or [− hearer-known] in the previous subsection is a somewhat idealized dichotomy (Hawkins, 1978; Lambrecht, 1994; Lyons, 1999; Prince, 1981). If we consider the newness of a NP as scalar rather than discrete, then the relative position of a NP along this scale might affect the use of articles. This hypothesis is outlined in (7).

(7) Hypothesis 4 (universal discourse knowledge): Speakers' use of articles will vary according to the relative givenness or newness of the NP in question.

Jarvis (2002, p. 388) pointed to the important difference between the status of a NP as the topic or comment of a clause, and its status as relatively new, known, or current in the ongoing discourse. He observed that Young (1996) defined topics as old information, thus conflating potentially distinct factors of clause-level and discourse-level information. I follow Jarvis in distinguishing between these two types of NP status: Hypothesis 1 will examine the topic (theme) or comment (rheme) status of a NP in its clause, whereas hypothesis 4 examines the overall discourse status of the NP.

METHOD

Participants and Interviews

Unlike many studies of indigenized nonnative English, which for various reasons have been limited to relatively proficient speakers, the present study encompasses a wide range of proficiency in English to reflect more accurately the complex nature of bilingual continua. The data were collected among first-generation adult Indian immigrants in California. All 12 individuals acquired English to varying degrees in India, emigrated as adults to the United States, and, for the most part, maintained their multilingual repertoires in the United States. At the time of the study, most of these subjects were employed in small commercial shops, and two were working in software companies, all within the San Francisco Bay Area. None of the individuals had taken English classes in the United States. Their increased informal contact with a native variety of English is an important new development in their linguistic environment, and they show socially motivated variation in their patterns of adoption of American phonological features in particular (see Sharma, 2005). However, for the present discussion, I focus on article use and do not specifically address the role of contact with American English.

Table 1 lists background information for all speakers; the speakers are ordered according to the results presented later in Table 2. The 12 speakers do not comprise an exact cross-classification of every external variable; however, because the goal was to broadly characterize features of the Indian English bilingual continuum, this shortcoming is balanced by the inclusion of a relatively diverse range of English proficiency levels in the group.

As a central interest of the present study is the characterization of the natural development and use of new nonnative dialect features, the data were collected through sociolinguistic interviews rather than through elicitation tasks. The interviews, ranging in duration from 1 to 2 hours, aimed at eliciting naturalistic speech data, personal demographic information, personal experiences and narratives, and information about the speakers' attitudes toward language use, dialects, and cultural contact. Because variation in article use in relation to discourse structure was examined, the relatively informal nature

Table 1. Participants' social characteristics

Characteristics	Participants											
	KD	SK	CK	RS	RR	KP	KK	GV	RT	KB	SS	NT
Education	0	0	0	1	1	1	2	2	2	2	2	2
Functional role	0	0	0	1	1	2	2	1	2	3	3	3
Years in the United States	2	18	17	2	17	25	39	0.5	0.5	40	0.7	2
Age	34	38	67	26	48	54	62	35	29	67	23	24
Sex	M	F	M	M	M	F	M	M	M	M	M	M
L1[a]	Guj	Pun	Guj	Tam	Guj	Guj	Pun	Kan	Hin	Hin	Hin	Hin

Note. Education: 0 = no formal English; 1 = English in higher education; 2 = primarily English education; functional role: 0 = minimally at work; 1 = regularly at work; 2 = work and some friendships/younger relatives; 3 = work, friendships, home; L1: Guj = Gujarati; Pun = Punjabi; Tam = Tamil; Kan = Kannada; Hin = Hindi/Urdu.
[a]Most individuals speak Hindi in addition to their L1.

of speakers' responses was advantageous. This stands in contrast to many studies of article use that have relied on elicitation tasks or grammaticality judgments (Agnihotri, Khanna, & Mukherjee, 1994; Goto Butler, 2002; Ionin et al., 2003; Kachru, 2003). Experimental tasks of this type are more controllable and can thus extract data for a complete range of contrasting article contexts; however, they almost invariably place speakers in highly formal and standard linguistic interactions, thus potentially eliciting maximally standard usage and possibly even distortions of natural usage due to the speaker's heightened linguistic self-consciousness in the test situation. Although these types of controlled elicitation tests might be appropriate for pedagogically motivated inquiry, they can skew attempts to investigate speakers' natural, unmonitored language use.

Implicational Profile of Participants' Grammars

As part of the description of the participants, a brief quantitative profile of their rates of use of certain grammatical features is included here, simply to indicate the salience of nonstandard article use among such speakers and to motivate the choice of this particular aspect as the focus in this study. Kachru (1965, pp. 393–396) developed the concept of a "cline of bilingualism" in order to introduce multidimensionality to the study of indigenized varieties of English, and although frequent reference has been made to this cline, few studies have sought direct empirical evidence for it in the natural speech of bilingual Indian English speakers. As implicational scaling was introduced into the study of Creoles (Bickerton, 1971; DeCamp, 1971) and of SLA (Andersen, 1978; Bayley, 1999; Pienemann, Johnston, & Meisel, 1993) in order to identify different dimensions of variation, it is well suited to a quantitative implementation of Kachru's notion (cf. Agnihotri et al., 1994; Ho & Platt, 1993). For the present study, this tool is of particular interest, as it can help establish

whether certain grammatical features—in this case, articles—are in fact more divergent in their usage patterns.

In an implicational relationship, the presence of a feature *x* implies the presence of *y*, but not vice versa. Thus, one may have a grammar that includes only the feature *y* or a grammar that includes both features *x* and *y*, but a grammar that exclusively includes the feature *x* is not predicted. In tabular form, this relationship translates into the claim that a value in one column will have higher values above and to its left, and the reverse will be true for lower values. Although implicational scaling is the subject of some debate in both Creole studies (see Rickford, 2002, for a summary) and SLA (Hudson, 1993; Huebner, 1983), I adopt it here merely as a preliminary organizing tool that highlights—rather than explains—individual differences across a speech continuum.

To contextualize articles within the grammars of these speakers, this scaling includes a range of grammatical features. Variable past-tense marking, copula use, and subject-verb agreement are common features of L2 English (Bayley, 1994; Dulay & Burt, 1974; Hawkins, 2001; Wolfram, 1985), as is variation in article use. The latter in particular has been noted as characteristic of Indian English (Agnihotri et al., 1994; Kachru, 1983; Platt et al., 1984; Williams, 1987).

Table 2 lists percentage rates of nonstandardness for each of these features for the 12 participants: past tense (e.g., *I stay(ed) in San Francisco last year*), copula (e.g., *He (is) crazy*), subject-verb agreement (e.g., *Our prices is (are) cheaper*), definite article (e.g., *(The) driver gave me directions*), nonspecific indefinite article (e.g., *I'm looking for (a) job*), and specific indefinite article (e.g., *I met (a) friend of yours*).

The scalability of an implicational distribution—calculated by dividing the number of correctly predicted values by the total number of values—represents the closeness of fit between the data and the predicted implicational model. In line with many previous studies, the high rates of scalability in both dimensions in Table 2 demonstrate that L2 speech is clearly structured despite considerable variation across speakers. Particularly in the horizontal dimension, only 8 values of 72 total violate the predicted ordering; furthermore, these violations generally remain within the expected range and do not diverge dramatically from the values in the neighboring cells.

Perhaps the more important finding in Table 2 is that the data can in fact only be arranged into two separate implicational scalings. This is because the variables conform fundamentally to strikingly different patterns of variation. The features in "Scale A" on the left side of Table 2—subject-verb agreement, past marking, and copula use—converge toward native English-like usage rapidly as one moves down the continuum, so that more proficient speakers such as RT, KB, SS, and NT do not show any indigenized usage of these features. In keeping with many learning trajectories, the distribution across these first three columns resembles an S-curve, whereby the majority of the values (34/36) in the distribution are restricted to lower (0–25%) and higher (70–100%) frequency ranges, and a minority (2/36) fall in a transitory intermediate frequency range (25–70%). It is worth noting that the actual ordering of these

Table 2. Percentage rates of nonstandard forms by speaker

	Scale A						Scale B					
	No past marking		No copula		Agreement mismatch		No definite article (evoked)		No indefinite article (nonspecific)		No indefinite article (specific)	
Speaker	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
KD	30	70	112	24	112	6	7	86	17	82	6	50
SK	66	70	63	21	102	16	4	50	19	84	12	50
CK	19	47	31	23	85	9	—	—	6	67	11	45
RS	169	28	109	16	127	3	33	55	38	58	32	31
RR	90	20	134	5	162	11	5	60	37	57	26	46
KP	78	22	106	5	130	8	9	78	19	47	13	23
KK	83	7	92	15	98	4	8	50	55	65	30	30
GV	76	3	130	2	112	0	13	62	57	51	28	21
RT	68	0	96	0	116	0	10	50	18	44	20	20
KB	86	0	93	0	122	0	5	40	28	29	53	21
SS	30	0	69	0	72	0	4	75	16	25	11	36
NT	131	0	112	0	134	0	4	0	70	9	32	0

Note. Horizontal scalability: 91.6% (columns 1–3), 86.1% (columns 4–6); vertical scalability: 86.1% (columns 1–3), 69.4% (columns 4–6). *n* values in all tables report the total number of relevant contexts in which nonstandard use of a variable might occur.

three interlanguage features is different from universal orders of acquisition proposed for English (cf. Dulay & Burt, 1974), and I make no claim that the order of interlanguage features apparent here is universal.

The considerable overlap between English and the L1s of these individuals with respect to these particular features might reinforce the ultimate acquisition of the native English system. There are five L1s represented by the 12 participants, and all five L1s mark past tense overtly, in some cases with more distinctions (e.g., in person, number, and gender) than English; all five also include some system of subject-verb agreement and, again, these systems are often more complex than English (e.g., person, number, and gender agreement as well as object agreement); and, finally, all five languages have copular verbs, although Tamil and Kannada permit a null copula in certain clauses (e.g., with nominal predicates).

By contrast, nonstandardness in article usage—"Scale B" on the right side of Table 2—is more systematically divergent and persists long after other L2 learning features have been acquired.[1] In some cases, this usage has come to be generalized to the extent that it appears in the speech of individuals who consider English one of their native languages (KB, SS, NT). Although gradually decreasing across the continuum of speakers, article omission does not

follow an S-curve trajectory; instead, it exhibits a flatter and higher distribution. This could either be taken as evidence of a new nonstandard article system beginning to be indigenously transmitted or as evidence of very late-stage SLA; regardless of the interpretation, we can minimally conclude that the two scalings in Table 2 offer empirical evidence of two distinct patterns of variation within the grammars of these stable bilinguals.

The article systems of the speakers' L1s are far more dissimilar to English than the L1 systems of tense, copulas, and agreement. At first glance, this might suggest that the greater divergence in article use derives from negative transfer from L1 systems. This paper aims to evaluate whether this is in fact the case and, if not, what the precise nature of the developing system is.[2]

Coding of Article Tokens

The four hypotheses presented earlier for article use are concerned with the omission of *the* and *a* in contexts that would require an overt article in standard English. Contexts in which standard English would not require an article were therefore not included in the coding. These contexts included indefinite noncount nouns (e.g., *I need furniture*) and all plurals—indefinite plurals do not occur with overt articles (e.g., *I teach children*), and generic and definite plurals have optional articles (e.g., *(The) prices went up*). Extraneous article insertion in contexts that standardly do not require an article (e.g., *We speak the Hindi*) occurred occasionally, but these uses are not discussed here. Following Huebner (1983), Tarone and Parrish (1988), and Young (1996), possessives, numerals, demonstratives, and quantifiers were excluded, mainly due to the lack of significant interaction between these determiners and *the* and *a*. Two sets of tokens were therefore coded for all singular count nouns: the indefinite article *a* (presence/absence) and the definite article *the* (presence/absence).[3]

The term *null* is used here to refer to the absence of *a* or *the* when either article would be standardly overt. Chesterman (1991) referred to the nonovert indefinite article (e.g., with mass and plural nouns) as the zero article, and the nonovert definite article (e.g., with proper nouns) as the null article. Adopting either of these meanings would be problematic, as both refer to standard article absence in English, and the article absence of interest here is nonstandard. Therefore, the term *null* is used in a different sense, simply to indicate the nonstandard absence of either article.

In one set of data, the dependent variable was the absence of the indefinite article, occurring in contexts where standard English would require the overt indefinite article *a*.[4] Examples of standard and nonstandard usage are presented in (8a) and (8b), respectively.

(8) a. I told my employees, if I am talking to *a customer* don't talk to me in Punjabi.
b. Then he thought, what about getting *Ø girl* [to marry] from India?

In the other set, the dependent variable was the absence of the definite article in contexts where standard English requires the overt definite article *the*. Standard and nonstandard examples of definite article use are given in (9a) and (9b), respectively.

(9) a. Berkeley is just like a little India, because of *the cosmopolitan area* and *the students*.
b. Here I'm not working in *Ø kitchen*—yeah, I'm in *Ø front desk*; I'm a manager.

In addition to coding for the presence or absence of the definite and indefinite articles, six independent variables were examined for each token. Most of the six independent internal (linguistic) variables coded derive from the hypotheses developed in the previous subsection. Other factors, such as grammatical function and modification, were included simply to include as wide a range of potential factors as possible.[5] The factors examined and the values assigned during coding were the following: (a) grammatical function (subject, VP object, PP object); (b) clausal topicality (theme, transition, rheme); (c) clause position (initial, medial, final); (d) specificity (specific, factive, nonspecific, generic); (e) modification (bare, modified, quantified); and (f) givenness (brand-new, new-anchored, unused, inferable, containing inferable, given). Standard examples of each of these internal variables, taken from the interviews, are given in (10)–(15), and a brief discussion of each variable precedes its examples. For each set of examples, the values assigned appear in parentheses.

Grammatical Function. Three types of grammatical function were distinguished; direct and indirect verbal objects were both coded under (10b). This factor is not discussed in detail, as it did not appear to be a strongly conditioning factor; its relative effect is noted later in Tables 8 and 9.

(10) a. *The future* is safe if you get all the things. (subject)
b. My wife owns *the store*. (VP object)
c. She's very involved in *the community* here. (PP object)

Clausal Topicality. Following Young (1996), three categories of clause topicality were coded: the topic of the clause (theme), the comment on the topic (rheme), and the linking or transitional term, if it occurred (Birner & Ward, 1998; Dik, 1978; Lambrecht, 1994). The transition tends not to be picked up as a new theme, once introduced, but rather serves a narrowing or specifying function. Degrees of discourse familiarity, as opposed to clausal topicality, were coded separately and are discussed in (15).

(11) a. If *the economic situation* would change I'm sure lot of Indian will go back. (theme)
b. This is *the major problem* I am facing here. (transition)
c. He had requested for *a Hindi teacher* (rheme)

Clause Position. Following Young (1996) again, three clause positions were coded. Initial clause position included all NPs that occur at the left edge of the main clause; final position included NPs at the right edge of the clause, regardless of grammatical function; and medial position included NPs that appeared in intermediate object or subject positions in complex clauses.

(12) a. *The other brother* was working in the factory. (initial)
b. So that's *the reason* so many Indian family are separated. (medial)
c. So six months I was there in *the kitchen*. (final)

Specificity. Rather than using a binary or ternary distinction for specificity and genericity, I followed Sankoff and Mazzie (1991) in adopting the additional category of "factive" for predicative NP constructions, as in (13b), which perform neither a strictly specific function nor an entirely nonspecific function, but rather mark category-membership. Sankoff and Mazzie treated factive as a variety of nonspecific; however, in the present data, factive NPs patterned almost identically to specific NPs in terms of null article use and were eventually grouped with specific NPs for the analysis. Most quantified NPs, such as *a couple of people* and *a lot of Indians*, were coded as nonspecific.

(13) a. There is *a association* that can help.[6] (specific)
b. My Mom is *a teacher*. (factive)
c. The parents will look for *a girl*. (nonspecific)
d. And you spend more time waiting for *the bus*. (generic)

Modification. The category of "quantified" in (14c) was fairly broadly interpreted and included ranking adjectives (e.g., *the first N*, *the last N*, *the best N*), quantifying phrases (e.g., *a lot of N*, *a few of N, the whole N*), and numeral modifiers (e.g., *the eleventh N*). The basic property that these categories share is that their modifier either uniquely isolates the referent or quantifies the referent set. By contrast, the category of "modified" included all other adjectivally modified NPs, in which the modifier simply restricts and specifies the potential real world referents.

(14) a. I worked as *a bartender*. (bare)
b. Bihar traditionally been *a poor state*. (modified)
c. It's a problem for *a lot of families*. (quantified)

Discourse Givenness. For the final independent variable, I relied on Prince's (1981) scale of assumed familiarity, further details of which are discussed later in relation to hypothesis 4. NPs were classified into six degrees of discourse familiarity—exemplified in (15a–15f)—based on a scale of increasing familiarity status that a NP might have at a given point in discourse. In (15a), the NP, *a Shiva temple*, is brand new in the discourse: It has not been

mentioned previously, and the hearer cannot anticipate it based on prior knowledge. In (15b), the new NP is anchored to a known discourse entity (*the restaurant*) and thus bears a slightly greater degree of familiarity to the hearer. In (15c) *the American government* is also new to the discourse but can be assumed by the speaker to be known to all participants in the discourse. In (15d), *the professors* is inferable information, as it has not been previously evoked but bears some prototypical relation to a familiar entity in the discourse and can therefore be anticipated by the hearer. (15e) contrasts with (15d) only in that the NP explicitly contains the evoked entity (*the work*) that provides the basis of the inference. Finally, in (15f), the second mention of *the company* is evoked: Its referent has already been mentioned recently in the discourse and can be assumed to have a familiar status for both speaker and hearer.

(15) a. We have *a Shiva temple*. (brand-new)
b. And it was *a policy of the restaurant* not to hire anybody with beard and long hair. (anchored-new)
c. And then one day, *the American government* was calling dentist, doctor, nurses over here. (unused)
d. They settle down over there so they have their own school. *The professors* were Indian and everything. (inferable)
e. Because of *the nature of the work*, I'm always busy. (containing inferable)
f. Some have a mentality of joining a start-up company. . . If *the company* goes to public, they become millionaires. (evoked)

These distinctions in relative discourse familiarity of an entity are shown in Figure 3. Based on an analysis of casual standard English speech, Prince (1981) argued that the subdivisions in Figure 3 actually form a scale of assumed familiarity, shown in (16).

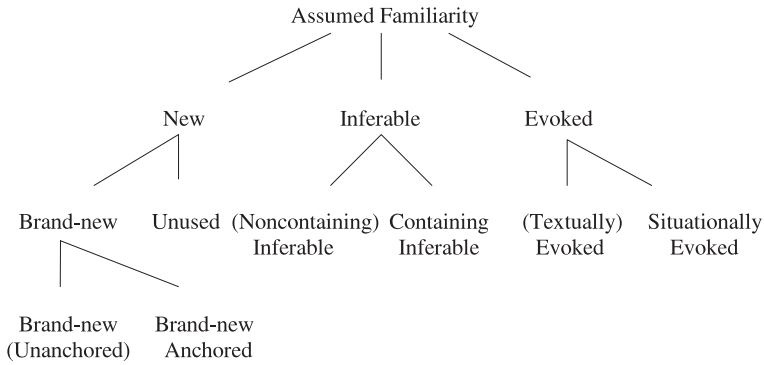

Figure 3. Taxonomy of assumed familiarity (Prince, 1981, p. 237).

(16) Evoked (E), Situationally evoked (E^S) > Unused (U) > Inferable (I) > Containing inferable (I^C) > Anchored-new (BN^A) > Brand-new (BN)

Prince used this scale to account for discourse-driven choices of linguistic forms, such as articles, pronouns, subjects, or modifying phrases. For instance, if a discourse entity (e.g., *John)* is known to both speakers but is unused, then stating that *John* (U) *bought a car* is felicitous, whereas *He* (E) *bought a car* and *A guy I know* (I^C) *bought a car* are both infelicitous: *He* implies that the referent is accessible to the hearer, and *a guy I know* implies that the hearer does not know the individual in question.

The process of determining the givenness status of NPs is not always straightforward. Prince (1981, p. 244) acknowledged that the determination of a NP as evoked, unused, or inferable can sometimes be difficult. For instance, a NP such as *my grandmother* could be considered inferable on the basis that individuals are known to have grandmothers, but it might also be considered unused information if the speaker and hearer both know the grandmother in question. To avoid these problems, I coded NPs as unused only if they were culturally recognized entities such as *the yellow pages*, *the Internet*, or *the U.S. government*. For this reason, the category of unused has relatively few tokens.

The distance in time and discourse also complicates the distinction between NPs that are evoked and those belonging to less familiar states. The status of a NP as evoked fades as the discourse progresses, until its evoked status can become unclear. Ariel (1990) used a measure of distance to determine which parts of the discourse were still recent in the interlocutors' minds; Lambrecht (1994) employed the notion of active (in the addressee's memory) and accessible (available but distant in discourse) discourse referents. In general, in the current study, it was possible to determine from the context whether a NP was active at a given time.

To remove other ambiguities from Prince's (1981) system, I was obliged to impose more explicit definitions on the categories of containing inferable and anchored-new as well. I interpreted the category of containing inferable to include a number of NP structures: NPs with complement phrase complements (e.g., *the salesclerk that we hired*), NPs with following locatives (e.g., *the biggest market in this area*), and ellipsis NPs (e.g., *the first son was born last year and the second is. . .*). The category of anchored-new included NPs with an evoked locative (e.g., *a man in the market*), NP heads of relative clauses (e.g., *a man I know*), prepositionally modified NPs (e.g., *a problem with the suppliers*), and modified NPs relying on prior discourse knowledge (e.g., *a newer store*). Equative NP predicates, as in (13b), were classified as anchored as well, because the identity of the referent is being indexed with a discourse-familiar subject; the only exceptions to this were when the construction was used with irrealis or negative meaning or when the NP was the predicate of an expletive subject. In such cases, the NP was classified as brand-new.

The lexical semantics of particular verbs occasionally presented coding challenges. For instance, if a speaker uses the verb *rent*, the likelihood of the fol-

lowing NP falling within the class of typically rented items such as real estate or vehicles is high. Nevertheless, the NP cannot be treated as strictly inferable, as it is not a canonical or predictable extension of something already evoked in the discourse (as is the case with "bus . . . driver"). Only NPs that were very strongly conditioned by the verb were coded as inferable; the rest were coded as new.

Finally, a wide range of NPs were disregarded in both datasets on the basis of criteria that I developed in the course of examining the data. As mentioned earlier, any plural contexts and contexts permitting optional article use in standard English were excluded. Proper names were excluded due to evidence of lexically specific article omission (e.g., *the United States*, *the Niagara Falls*). Articles within fixed constructions such as *most of the N*, *all of the N*, *such a N*, and in adverbial uses such as *a little* and *a lot* were also excluded, again due to the potential for lexically specific biases.

To evaluate each hypothesis across the continuum of speakers, the 12 individuals are arranged into three groups. Rather than dividing them evenly into three groups of four each, I followed the clustering of speakers that emerged in Tables 1 and 2. Thus, among the three speakers in group 1 (KD, SK, CK), English has the lowest functional and educational level in their repertoire; among the five speakers in group 2 (RS, RR, KP, KK, GV), it is used regularly as a L2; and among the four speakers in group 3 (RT, KB, SS, NT), English is often on equal footing with other L1s. These groupings ease the presentation of data, and in the discussion that follows, I focus on whether each hypothesis is confirmed for the group as a whole.

RESULTS

Evaluation of Hypotheses 1–3

Evaluation of Hypothesis I: Transfer of L1 Positional Marking of Discourse Status. Hypothesis 1 anticipates that initial themes and final rhemes will be the categories favoring nonuse of articles, as these are the unmarked contexts for old and new information, respectively. Table 3 shows that no statistically significant pattern emerges for either of these predictions.[7] In fact, when separate chi-square tests are conducted for each group, only group 1 has a significantly higher rate of article omission with initial themes, $\chi^2(1) = 4.01$, $p < .05$. This slight confirmation of the hypothesis is not repeated in group 1's omission of articles with final rhemes; in fact, their rate of omission of articles with final rhemes appears to run slightly counter to the prediction. Thus, although Table 3 shows that rhemes tend to cluster in final position and themes in initial position in terms of overall frequency, a link between these positions and article use is not evident.

Evaluation of Hypothesis 2: Transfer of L1 Article System. The first prediction of hypothesis 2 is that the L2 English system will imitate the L1

Table 3. Null article use according to clause position and topicality

	Initial theme NPs		Noninitial theme NPs		Final rheme NPs		Nonfinal rheme NPs	
Group	*n*	%	*n*	%	*n*	%	*n*	%
1	38	74	4	25	47	53	29	59
2	101	50	33	48	333	44	87	44
3	68	29	22	36	180	16	85	15
Total	207	47	59	42	560	36	201	34

Note. Initial theme NPs versus noninitial theme NPs, $\chi^2(1) = 0.46$, $p = .4976$; final rheme NPs versus nonfinal rheme NPs, $\chi^2(1) = 0.23$, $p = .6315$.

system by marking only specific indefinites with an overt article. This is supported by the data in Table 4: A statistically significant difference across groups is found such that specific indefinite NPs—which would have an overt article in the L1s of the speakers—have rates of null marking that remain below 50% even for group 1. On the other hand, nonspecific indefinites, which are null marked in the L1s, are null marked at much higher rates.

However, the common claim that Indian speakers use *one* in place of *a* with specific indefinite NPs was not supported. Although individuals showed a noticeably higher rate of use of *one* in standard (emphatic, specific) reference than native English speakers might have, very few instances were clearly nonstandard, most of which are given in (17).

(17) a. Some entertainment workers are there. They are work in theatres for *one* dance.
b. They opened *one* Udipi Palace in Seattle too.
c. Like, I had *one* big joke, you know, on that part. . .
d. There was *one* Britisher come over there.
e. First of all, the Indians have *one* distinct accent to begin with.
f. I went to Home Depot. There's *one* girl, no, she's billing everything and she looked my t-shirt. . . .

Table 4. Null article use with definite and indefinite NPs

	Indefinite specific NPs		Indefinite nonspecific NPs		Definite NPs	
Group	*n*	%	*n*	%	*n*	%
1	29	48	42	81	83	46
2	129	31	206	57	343	38
3	116	16	132	20	209	18
Total	274	27	380	47	635	33

Note. Specific indefinite versus nonspecific indefinite NPs, χ^2 (1) = 26.8, $p < .001$.

Although Platt et al. (1984, p. 56) listed this as an important transfer effect in their example given in (3a), the far more robust quantitative pattern in the present data is the indirect transfer of function seen in Table 4, whereby *a* is restricted to a specificity-marking function. The use of *one* might be more noticeable in Indian English and thus cited more often in nonquantitative studies, due to its greater salience rather than its actual quantitative frequency.[8]

The second prediction of hypothesis 2—namely, that the definite article will be absent because the L1 does not have one—is not clearly supported in the data reported in Table 4. The prediction anticipates that at least group 1 should have very high rates of nonuse of definite articles; however, even for this group, the rate of null definite articles does not exceed 50%. Thus, hypothesis 2 is confirmed only in the part of the grammatical subsystem in which the L1 has an overt form. Where a gap occurs in the L1 (i.e., no definite article), we do not find a matching absence of use in the L2 English grammar; instead, there appears to be a nearly even rate of overt and null use of the definite article.

Evaluation of Hypothesis 3: Universal Tripartite Article System. Hypothesis 3 predicts a three-way distinction based on universal contrasts, namely that there will be distinct forms used for [+ specific reference, + hearer-known] and for [+ specific reference, − hearer-known] and that there will be a zero form for [− specific reference, ± hearer-known]. Table 5 therefore presents rates of overt usage for the first two categories and null use for the latter; thus, according to hypothesis 3, all rates in Table 5 should be high, particularly for group 1.

What we find is that although group 3 approaches the predicted high rates for the first two categories, because standard English marks the distinction between [+ hearer-known] and [− hearer-known], this is not the system that the less proficient groups universally adopt. The only potential support for an emergent universal tripartite system is the high rate of null article use with nonspecifics and generics that is apparent in group 1. Groups 2 and 3 decline in their rates of null use as they move toward the English sys-

Table 5. Article use according to predicted universal tripartite system

	Overt *the* with definite NPs		Overt *a* with indefinite NPs		Null article with nonspecific/generic NPs	
Group	*n*	% overt	*n*	% overt	*n*	% null
1	72	56	29	52	53	75
2	248	60	129	69	301	50
3	193	81	116	84	148	19
Total	513	67	274	73	502	43

tem, which does not have null-marked nonspecifics. However, even this single piece of support for the hypothesis can be equally well accounted for by hypothesis 2, which explains this pattern for group 1 speakers in terms of the re-creation of their L1 specific/nonspecific distinction in their L2 article use.

In sum, none of the first three hypotheses is completely confirmed. Hypothesis 1 was partly confirmed insofar as null definite articles were used with clause-initial themes; however, as this hypothesis treats discourse status in terms of salient contrasts in clause position, a partial confirmation of the hypothesis is insufficient. Hypothesis 2 was also partly supported within the domain of specificity marking, but not in terms of null use of definite articles. Hypothesis 3 only found support in the area of null marking of nonspecifics, a pattern that can also be explained by hypothesis 2.

Taken together, the main finding thus far is that a transfer effect can be observed within the domain of specificity marking; variation in use of the definite article remains largely unexplained. An important implication of this difference is that transfer effects may be stronger in the part of the grammar where the L1s have an overt form—namely, a specific indefinite article. In the next subsection, I explore the final hypothesis more closely, using a more refined model of discourse universals.

Evaluation of Hypothesis 4: Discourse Familiarity

A fundamental problem with the approach to definiteness in the preceding discussion has been its treatment as a binary opposition, a critique that has been made elsewhere (see Chesterman, 1991, p. 39, for a summary). In this subsection, I use a more fine-grained model of givenness to consider the exploitation of discourse knowledge by speakers in their attempts to create order in their L2 grammars. Sankoff (1983) cited the crucial role of agentive exploitation of linguistic contrast for discourse purposes by nonstandard users of a language:

> It has been proposed that creolization involves people generating linguistic rules for which they have no evidence in the input. I suggest that what people do is reanalyze 'grammatical' input that is generated 'discursively'—as the insertion of *pas* redundantly and emphatically in negative sentences in French was originally a discourse strategy that later became grammaticalized. The same sorts of strategies arise over and over in language, as people hit on the same solutions to their expressive problems, but rarely do these innovations become institutionalized—both first and L2 learners must eventually conform to the preexisting norms of the languages they are learning. The genesis of both pidgins and creoles has taken place under conditions where the innovative strategies do get more of a chance to survive, because they are not competing in the same way with existing rules. (p. 245)

Indigenized nonnative varieties share with creoles this widespread absence of native target norms, permitting a stabilization of discursive solutions.

Modeling Givenness. Models of definiteness generally share the feature of *identifiability* (Lambrecht, 1994; Lyons, 1999). Articles frequently begin their life as demonstratives or topic markers—highly discursive constructs—and might then gradually grammaticalize to cover a wider or narrower range of contexts (Givón, 1984; Greenberg, 1978; Lyons). For instance, although English has not extended the definite article to proper nouns or plural generics, Greek uses it with the former and French with the latter (Trenkic, 2002, p. 109). It is possible that these discursive origins partly resurface through the process by which nonnative speakers often undo certain grammaticalizations and move toward transparent, analytic morphology.

Hypothesis 4 explores whether there is a universal discourse effect governing article use in Indian English on the basis of this function of identifiability. As identifiability can be thought of as scalar, the models of particular interest are those that distinguish among several types or degrees of definiteness or givenness, such as Givón (1984), Hawkins (1978), and Prince (1981); the suitability of these more fine-grained taxonomies for the study of nonstandard article variation has been noted elsewhere as well (Kachru, 2003; Sankoff & Mazzie, 1991). Prince's (1981) scale of assumed familiarity—introduced earlier in relation to Figure 3—has a number of advantages for the present analysis: It focuses on discourse entities (or NPs) and so it can be applied directly to the question of article use; it appeals to universal discourse principles and thus represents—in a L2 situation—an alternative potential universal influence; and it makes implicational predictions for how discourse entities at different levels of the scale will relate to types of linguistic expression, allowing us to move beyond binary contrasts. In some respects, Hawkins' (1978) taxonomy allowed more detail than Prince's model; for instance, his distinctions among associative anaphoric use (*a book. . . the author*), larger situation use (*in a village. . . the church*), and explanatory modifier use (*I remember the beginning of the war very well. . .*) are all subsumed under Prince's category of inferables. However, his taxonomy does not order the various types that he identified with respect to one another, a feature of particular interest to the present study.

Article Use According to Relative Givenness. Standard examples of each type of discourse reference in Prince's model were given in (15) and nonstandard null article examples from the present dataset are given in (18):

(18)	a. Brand-new	(BN)	We decided to rent *Ø apartment.*
	b. Anchored-new	(BN^A)	They speak *Ø different kind of English.*
	c. Unused	(U)	You can browse through *Ø internet.*
	d. Inferable	(I)	If I'm marrying . . . what about getting *Ø girl* from India?
	e. Containing inferable	(I^C)	From *Ø first year of birth* until. . .
	f. Evoked	(E)	In tenth the medium was English. After that I took science. *Ø medium* was English.

Table 6. Null article use according to familiarity status of NP

	BN		BN^A		I^C		I		E	
Group	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
1	11	36	15	53	13	46	50	40	11	73
2	29	7	87	34	62	35	126	37	57	60
3	42	5	73	22	74	8	81	17	23	43
Total	82	10	175	31	149	23	257	32	91	57

Note. $\chi^2(4) = 51.52$, $p < .001$.

Although the standard English article system can be described by Prince's (1981) model to the extent that definite articles are associated with E, I^C, I, and U and indefinite articles are mainly associated with BN^A and BN, null realization of articles does not relate to this scale.[9] Indian English, however, appears to extend the scale of assumed familiarity to include the option of null use as well. Rates of article omission relative to the discourse givenness of a NP are shown in Table 6. To temporarily factor out the L1 effect on nonspecific articles (see Table 4), Table 6 only includes definite and indefinite NPs with specific reference.

As Figure 4 also shows, omission of the article—whether *the* or *a*—increases overall relative to the familiarity status of the NP. This statistically significant pattern holds within groups as well: Group 1 has predictably higher rates overall, followed next by group 2, and, finally, the rates of group 3—the most proficient speakers—are the lowest. In all three groups, brand-new NPs have

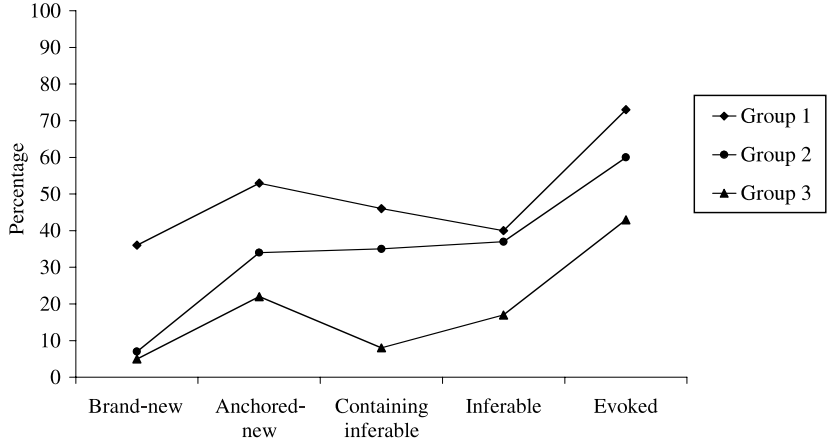

Figure 4. Null articles according to givenness (*n* values are given in Table 6).

the lowest rates of article omission, and evoked NPs have the highest rates of article omission. The intermediate categories of Prince's (1981) model—anchored-new, inferable, and containing inferable—show intermediate and equivalent levels of article omission.

Overall, Figure 4 indicates that a systematic use of overt articles mainly for purposes of discourse disambiguation might be operative for these Indian English speakers. The pattern shows that Prince's (1981) scale—which operates on other domains of standard English syntax—can also extend to new grammatical domains in nonnative discourse. However, it should be noted that not all contrasts in Prince's scale appear to be equally salient to the speakers. The contrasts of new/anchored-new, new/inferable, and inferable/evoked show greater differences than inferable/containing inferable or anchored-new/inferable. Also, the category of unused was highly variable, sometimes patterning like a new category and, at other times, resembling the rates for inferable NP types; as the number of unused NP tokens was low, it was excluded from Table 6.

Let us consider an example of this new pragmatic article use in detail. The extract in (19) is taken from Mishra (1982)—a study of prosody in Indian English—and although the study does not address article use, the actual speech extract used includes a rather striking demonstration of the alternations in marking of relative discourse givenness by Indian speakers. The fact that this extended illustration comes from the natural speech of an Indian speaker recorded for a completely different purpose lends support to the idea that this usage is widespread among Indian speakers. After the full extract, individual examples are highlighted in (19a–f).

(19) When I had completed the training, ten day training at the language school, and you know that what happened there, there was another week for the vacation. And during that vacation I contacted the union and union person contacted his representative at the school. And that representative contacted the headmaster and headmaster had contacted the authority. But before that instance in the morning, first day of the term, I had met him and told him that I'm worried.

(a)	When I had completed *the training*,	I
	Ø ten day training at the language school	E
(b)	and you know that what happened there	
	there was another week for *the vacation*	I
	and during *that vacation*	E
(c)	I contacted *the union*	I
	and *Ø union* person	I^C
(d)	contacted *his representative* at the school	I^C
	and *that representative*	E
(e)	contacted *the headmaster*	I
	and Ø headmaster had contacted the authority	E
(f)	but before that instance in *the morning*,	I
	Ø first day of the term, I had met him and told him that I'm worried.	E

This extract reflects the characteristic alternation of inferable and evoked NPs, a far more common pattern in natural dialogue than the alternation between brand-new and evoked, according to Prince (1981). What is notewor-

thy here is the exceptionless alternation in article choice, whereby overt *the* or another definite determiner such as *his* is used with inferables (I and I^C), but either *that* or *Ø* is used for subsequent evoked (E) references to the same entity. The shorter examples from my own data, given in (20), show a similar alternation such that a new or inferable referent is preceded by the requisite article, whereas a following inferable or evoked reference to the previous referent omits the article.

(20) a. I don't like *the climate. Ø climate* too much cold. [KD]
b. It was *a very small town* I used to live. So you had to migrate to *Ø other town* for, like, after the seventh grade. [KP]
c. Somehow he thought ki [that] if I'm marrying *a girl*, she should support him that I should continue enjoying my music. And then he thought, what about getting *Ø girl* from India? [KK]
d. So in tenth, *the medium* English. After that I took science. *Ø medium* is English. [RT]
e. They refer for *the credit history*. Where they don't have *Ø credit history*, there I'm not getting *the credit card*. So now I have to go for *Ø secured credit card*. [GV]

Although these examples illustrate the importance of evoked status in article omission, they also point to other potential factors. For instance, the last NP in (20e) omits the article with the addition of the modifier *secured*, and (20a) and (20d) omit articles when the NP moves to a subject position. To evaluate whether it is the NP's status as evoked—more than any other factor—that is conditioning article omission, multivariate analysis is needed. Following a brief discussion of one remaining variable (NP modification), a summary of the relative influence of all factors is presented.

Further Support for Identifiability: NP Modification and Article Omission. A final internal variable that showed a striking influence on article realization is type of modification in the NP. In fact, this pattern lends support to the preceding finding that NPs with unambiguous discourse reference can drop their article more readily. If clarity of discourse reference and economy are major considerations for these speakers' decision to use or omit an article, then a modifier or a quantifier—which serves the function of specifying the referent from within a possible range—should render the article even more redundant and therefore more omissible. This is precisely the pattern found in the data.

Examples of quantified NPs without articles from the present data are given in (21); this influence of modifiers and quantifiers was observable in (19) as well.

(21) a. I studied in Gujarati, but *Ø second language* was English. [RR]
b. So it's more easier for new generation to understand *Ø whole concept* of the new technology and everything. [RR]
c. And now he has one son and *Ø second* is on the way. [KK]

Table 7. Null article use according to type of modification in NP

	Bare NPs		Modified NPs		Quantified NPs	
Group	*n*	%	*n*	%	*n*	%
1	83	43	45	58	26	92
2	379	29	186	49	113	78
3	250	16	124	17	83	28
Total	712	26	355	39	222	61

Note. $\chi^2(2) = 92.00$, $p < .001$.

d. In fact you find *Ø lot of Andhra guys* here rather than Bangalorians. [GV]
e. So they went to India every year from day one. From *Ø first year of birth* till. . . [KB]

The statistically significant distribution in Table 7 and the selected examples in (21) indicate that whereas bare nouns require overt articles, modified nouns are more likely to be associated with omission of the article, and quantified nouns actually favor null articles. If modification restricts the potential real-world referents of a NP token (e.g., *the cold day*) and quantification is understood as uniquely identifying the referents (e.g., *the first day*), then we can argue that exactly the same principle of disambiguation that was found for relative givenness lies behind this pattern as well.

In a quite different, more pedagogically motivated analysis of article use among Indian speakers, Agnihotri et al. (1994, pp. 185, 188) also found that articles are more frequently omitted when adjectives are present and when a NP includes a superlative quantifier. Although no clear explanation is given in their discussion, their results might reflect the pattern in Table 7. Thus, two of the most robust factors for omission of articles in the present data—NP modification and discourse givenness—both point to a single discourse pragmatic system of article use primarily for the purpose of disambiguation.

Relative Significance of All Factors

This discussion of article use among Indian English speakers has pointed to the particular importance of L1 transfer of specificity marking, on the one hand, and of discourse identifiability, on the other. A multivariate analysis is presented in this closing subsection in order to further bolster the claim that these are the most influential of a range of factors. The statistical software package VARBRUL (Rousseau & Sankoff, 1978), used widely in sociolinguistics, has been introduced into the study of SLA (Bayley & Preston, 1996; Berdan, 1996; Young, 1996) and has permitted a more careful consideration of

classroom SLA, especially in cases of societal bilingualism (e.g., Mougeon & Beniak, 1996).

Although the separation of speakers into three groups highlighted important proficiency-based differences earlier, for the VARBRUL analysis the entire group is treated as a single speech community. I separate null use of definite and indefinite articles in this analysis, as there might be instances in which a factor is relevant for one article but not the other, in which case combining the results for both articles would obscure the partial pattern. Table 8 presents the results for definite article use, and Table 9 presents the results for indefinite article use. Because the stepwise regression takes into account the relative effect of all the other factors, the best indicator of the strength of a factor

Table 8. VARBRUL results for definite article use (internal variables)

Variable	Tokens (*n*)	Null (*n*)	Null (%)	VARBRUL weight
Givenness				
Evoked	102	58	57	.779
Unused	54	18	33	.589
Inferable	323	98	30	.449
Cont. inferable	156	34	22	.371
Range				408
Modification				
Quantified	100	51	51	.774
Modified	127	58	46	.654
Bare	408	99	24	.378
Range				396
Clause position				
Initial	220	97	44	.683
Medial	146	33	23	.471
Final	269	78	29	.362
Range				321
Clausal topicality				
Rheme	263	83	32	.597
Theme	210	87	41	.453
Transition	162	38	23	.403
Range				194
Grammatical function				
Subject	203	89	44	[.582]
Verbal object	182	47	26	[.466]
Prep. object	250	72	29	[.457]
Range				125
Specificity				
Generic	122	41	34	[.546]
Specific	513	167	33	[.489]
Range				57

Note. Input value: .275; significance threshold: .05; values in brackets are nonsignificant.

Table 9. VARBRUL results for indefinite article use (internal variables)

Variable	Tokens (*n*)	Null (*n*)	Null (%)	VARBRUL weight
Specificity				
Generic	14	13	93	.971
Nonspecific	366	164	45	.564
Factive	183	56	31	.397
Specific	91	17	19	.325
Range				646
Modification				
Quantified	122	84	69	.818
Modified	228	80	35	.489
Bare	304	86	28	.361
Range				457
Givenness				
Inferable	68	46	68	.804
Anchored-new	313	135	43	.544
Brand-new	273	69	25	.365
Range				439
Clausal topicality				
Theme	56	36	64	.735
Rheme	498	185	37	.502
Transition	100	29	29	.354
Range				381
Grammatical function				
Subject	29	20	69	[.633]
Verbal object	504	191	38	[.506]
Prep. object	121	39	32	[.443]
Range				190
Clause position				
Final	400	156	39	[.527]
Medial	213	70	33	[.472]
Initial	41	24	59%	[.383]
Range				144

Note. Input value: .345; significance threshold: .05; values in brackets are nonsignificant.

is the column containing the VARBRUL weights in each table. The relative magnitude of the weight relates to the relative strength of the effect of that factor on the dependent variable. A weighting below .5 indicates that the factor in question favors overt use of the article, whereas a weight greater than .5 means that the given factor favors omission of the article; a weight close to .5 means that the factor has little or no effect. Statistical significance or lack thereof, as determined by the regression analysis, is also indicated in this column through bracketing of values that were not found to be statistically significant by VARBRUL. The input value—or input probability—listed at the bottom of each table refers to the average probability that a rule will apply; thus, for instance

in Table 9, a weighting of more than .5 indicates that null article will occur with a greater frequency than 0.345. The independent factors in Tables 8 and 9 are ordered according to their relative influence on the dependent variable.

In Table 8, the factors most strongly favoring null use of definite articles are discourse givenness and modification. Both of these factors also exhibit a scalar effect; quantification favors null marking the most, modification slightly less so, and absence of all modification the least of all. Similarly, evoked NPs strongly condition null marking, inferable NPs slightly less so, and brand-new NPs least of all. In Table 8, grammatical function and specificity were not found to have a statistically significant impact on the use of definite articles. In other words, these two factors were rejected by the multivariate analysis for definite article use. The latter is not surprising, as specificity distinctions fall primarily within the domain of indefinite entities. Table 9 shows that two of the three strongest factors in indefinite article use are the same as for definite article use, namely, discourse givenness and modification. The scalar patterns for these two variables are also similar to the patterns found for definite article use. In addition to these two factors, Table 9 shows that specificity is a strong factor, reflecting the L1 transfer effect noted under hypothesis 2. The factors rejected as nonsignificant in Table 9 are grammatical function and clause position. The fact that clause position only appears to matter with definite articles also reflects the earlier finding with respect to hypothesis 1 that clause-initial themes favored null marking.

The factors that influence use of definite and indefinite articles thus form a relatively coherent set, and the statistical regression shows that the strongest factors are indeed givenness, modification, and specificity, outweighing such factors as clause position and grammatical function. This type of statistical analysis can thus settle issues that have remained unclear in previous studies—for instance, the relative importance of topichood as opposed to subjecthood when high rates of clause-initial null articles are apparent (Huebner, 1983; Parrish, 1987).

DISCUSSION AND CONCLUSION

This study has aimed to advance our understanding of new nonnative dialects of English through a close analysis of factors in grammatical divergence. The initial implicational scaling found that certain features diverge much more dramatically than others, and the markedly different article system seems to be sensitive not only to specificity but also to the relative discourse familiarity of the NP. These findings promise a more complete account of why Indian speakers vary in their use of articles, as the study has isolated not only the more commonly cited L1 effects but also new, possibly universal, pragmatic functions.

Klein (1980), discussing variation in the use of synthetic and analytic forms of progressive aspect among Spanish-English bilinguals, observed that "the

actual choice between the more precise and the less precise alternatives. . . should be determined by pragmatic strategies based on relative need for precision" (p. 77). Klein and Perdue (1992) also argued that individuals with different L1s have a basic variety at one stage of learning "which seems to represent a natural equilibrium between the various phrasal, semantic, and pragmatic constraints" (p. 311). This integral role of pragmatics also echoes Tarone and Parrish's (1988) finding that articles with referential definites were employed with greater nativelike accuracy in narrative genres due to the greater communicative burden of precise and efficient reference.

The capacity for more given entities to drop their articles has been noted in a number of other studies, with varying degrees of explanation (Agnihotri et al., 1994; Bruyn, 1995; Jarvis, 2002; Sankoff & Mazzie, 1991). Jarvis explicitly found this constraint to be active.

> In the Finns' data, on the other hand, the use of Ø probably does not represent a simplified register as much as it represents the L1 Finnish convention of avoiding (what Finns perceive to be) redundant markers of definiteness and indefiniteness when these properties of an NP are already salient in a given discourse context. (p. 416)

He cited Givón's description of the relatively greater burden on processing of newer information, a factor that appears to be strongly reflected in the present data as well: "more continuous, predictable, nondisruptive topics will be marked by *less marking material*; while less continuous, unpredictable/surprising, or disruptive topics will be marked by *more marking material*" (Givón, 1984, p. 126).

Studies of Creoles have also found null marking with definites, despite the bioprogram prediction that null marking will be associated with generics. Bruyn (1995, p. 73) noted in passing that—counter to Bickerton's (1981) predictions—null-marked NPs in Sranan are not always generic and can also be definite. Sankoff and Mazzie (1991, p. 7) also found definite NPs that were null marked in Tok Pisin, and they cited Corne's (1977) study of Seychelles Creole, in which this author noted that definite nouns might be unmarked "when no ambiguity is possible" (p. 14).[10] Sankoff and Mazzie furthermore observed that both Bickerton and Givón (1984) had instances of null-marked definites in their own data but explicitly chose to exclude them. Givón discounted these instances as referring to characters peripheral to the main story; Bickerton (1977), using somewhat circular reasoning, chose to discard such cases because they did not fit his definition: "A total of 260 zero articles are found in the texts under consideration. 24 of these—almost one-tenth—may immediately be discarded, since they refer to things which have definite reference" (p. 235). This is not to say that null marking cannot be associated with generics, but given the present data and the considerable range of studies that have made passing reference to the occurrence of null article use with given NPs, the availability of a competing discourse principle for the omission of articles cannot be ignored.

However, as Jarvis (2002, p. 414) noted, the marking of broad discourse distinctions such as these in L2 English is often triggered by the existence of differences between the L1 and L2 systems. Thus, although discourse knowledge plays a role in the restructuring of these Indian speakers' English grammars, these forces might only intervene due to the initial disruption of significant mismatches between their L1s and English and the resulting need to settle on some principled basis for using article forms.

To conclude, this study has shown that systematic divergence in a stable nonnative variety can indeed be identified in quantitative terms. Furthermore, the results indicate that, rather than acting as opposing forces, language transfer and universals might enter into complementary partnerships. In the present case, the existence of an overt specific article in individuals' L1s appears to have a strong influence on their use of the English indefinite article, but their L1 lack of a definite article invites the imposition of other discursively available pragmatic principles for the use of English articles. As a result, the influence of factors that seemed on the surface to be unrelated—namely discourse familiarity and NP modification—can also be subsumed under a single pragmatic explanation.

(*Received 18 April 2005*)

NOTES

1. Only the rates of null usage for evoked definite articles are listed in Table 2—in other words, the use of definite articles with NPs that have already been mentioned in the discourse. As discussed later in this study, Prince (1981) treated this as the most "given" status for a NP, and it is this category that exhibits the highest rates of article omission.

2. A comparison with Table 1 shows that the external (social) factors that correlate most closely with these rates of nonstandardness are educational background and functional use of English, thus approximating a proficiency-based continuum (see Sharma, 2005).

3. All coding was done by the author. To check for reliability, a random selection of 100 NPs were coded independently by a trained researcher. Before discussion between coders, the average interrater reliability was 97.9% for standardness judgments on copula, agreement, past tense, and articles, and 93.25% for the six internal factors coded for article tokens.

4. For the purposes of this study, standard British and American English are treated as generally equivalent systems. This is not always the case, as in dialectal variants such as *in (the) hospital*. Such cases would have been omitted from the dataset, but none were encountered during coding.

5. External (social) factors were also coded but are discussed separately in Sharma (2005).

6. Variation in *a* and *an* occurs in the data but is not included as a variable. *a* is occasionally generalized to positions that standardly require *an*.

7. The total number of tokens in Tables 4, 5, and 7 is 1289. This total is slightly lower in Tables 3 and 6. This is because in Table 3 the category of "transition" is omitted (only "theme" and "rheme" are relevant), and in Table 6 only specific NPs are counted, and the category of "unused" is omitted due to low *n* values.

8. Another adaptation of an English form to a substrate function occurs specifically in the speech of the two South Indian speakers. Kannada and Tamil both allow demonstrative and quantifier forms to be used as indefinite articles, and the two speakers of these languages frequently used *some* and *this* as specific indefinite articles and *that* as a definite article.

9. Discourse-based article omission is not completely absent in native varieties, but is highly restricted to certain types of subject: Definite subject articles (*Ø Last film I watched was Star Wars*) and indefinite subject articles (*Ø Man walks into a bar. . .*) are occasionally omitted. Articles with

objects are not omissible in standard English casual speech, whether indefinite (**A man wants Ø gift for Ø 3-year-old kid*) or definite (**I told Ø man that the shop was closed*).

10. Sankoff and Mazzie (1991) also employed Prince's (1981) model in order to access the role of discourse status distinctions in article choice in Tok Pisin. However, their study is not directly comparable to the present one, as their focus was on the specialization of different forms (e.g., *dispela*, *ia*, *wanpela*) for different functions, mirroring a pattern shared by the Austronesian substrate languages.

REFERENCES

Agnihotri, R. K., Khanna, A. L., & Mukherjee, A. (1994). The use of articles in Indian English: Errors and pedagogical implications. In R. K. Agnihotri & A. L. Khanna (Eds.), *L2 acquisition: Sociocultural and linguistic aspects of English in India* (pp. 178–198). New Delhi: Sage.

Andersen, R. (1978). An implicational model for second language research. *Language Learning*, *28*, 221–282.

Ariel, M. (1990). *Accessing NP antecedents*. London: Routledge.

Bamgbose, A. (1998). Torn between the norms: Innovations in world Englishes. *World Englishes*, *17*, 1–14.

Bayley, R. (1994). Interlanguage variation and the quantitative paradigm: Past tense marking in Chinese-English. In E. Tarone, S. M. Gass, & A. Cohen (Eds.), *Research methodology in second language acquisition* (pp. 157–181). Mahwah, NJ: Erlbaum.

Bayley, R. (1999). The primacy of aspect hypothesis revisited: Evidence from language shift. *Southwest Journal of Linguistics*, *18*, 1–22.

Bayley, R., & Preston, D. (Eds.). (1996). *Second language acquisition and linguistic variation*. Amsterdam: Benjamins.

Berdan, R. (1996). Disentangling language acquisition from language variation. In R. Bayley & D. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 203–245). Amsterdam: Benjamins.

Bhatia, T. (1993). *Punjabi: A cognitive-descriptive grammar*. London: Routledge.

Bickerton, D. (1971). Inherent variability and variable rules. *Foundations of Language*, *7*, 457–492.

Bickerton, D. (1977). Pidginization and creolization: Language acquisition and language universals. In A. Valdman (Ed.), *Pidgin and creole linguistics* (pp. 49–69). Bloomington: Indiana University Press.

Bickerton, D. (1981). *Roots of language*. Ann Arbor, MI: Karoma.

Birner, B., & Ward, G. (1998). *Information status and noncanonical word order in English*. Amsterdam: Benjamins.

Bruyn, A. (1995). *Grammaticalization in creoles: The development of determiners and relative clauses in Sranan*. Amsterdam: Instituut voor Functioneel Onderzoek van Taal en Taalgebruik (IFOTT).

Cardona, G. (1965). *A Gujarati reference grammar*. Philadelphia: University of Pennsylvania Press.

Chaudron, C., & Parker, K. (1990). Discourse markedness and structural markedness: The acquisition of English noun phrases. *Studies in Second Language Acquisition*, *12*, 43–63.

Cheshire, J. (Ed.). (1991). *English around the world*. New York: Cambridge University Press.

Chesterman, A. (1991). *On definiteness: A study with special reference to English and Finnish*. New York: Cambridge University Press.

Corne, C. (1977). *Seychelles creole grammar: Elements for Indian Ocean proto-creole reconstruction*. Tübingen, Germany: TBL-Verlag Narr.

Dik, S. (1978). *Functional grammar*. Amsterdam: North-Holland.

DeCamp, D. (1971). Toward a generative analysis of a post-creole speech continuum. In D. Hymes (Ed.), *Pidginization and creolization of languages* (pp. 349–370). New York: Cambridge University Press.

Dulay, H. C., & Burt, M. K. (1974). Errors and strategies in child second language acquisition. *TESOL Quarterly*, *8*, 129–136.

Fodor J. D., & Sag, I. (1982). Referential and quantificational indefinites. *Linguistics and Philosophy*, *5*, 355–398.

Givón, T. (1984). The pragmatics of referentiality. In D. Schiffrin (Ed.), *Georgetown University round table on language and linguistics* (pp. 120–138). Washington, DC: Georgetown University Press.

Goto Butler, Y. (2002). Second language learners' theories on the use of English articles. *Studies in Second Language Acquisition*, *24*, 451–480.

Greenberg, J. (1978). *Universals of human language*. Stanford, CA: Stanford University Press.
Halliday, M. A. K. (1967). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics*, *3*, 199–244.
Hawkins, J. (1978). *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Croom Helm.
Hawkins, R. (2001). *Second language syntax: A generative perspective*. Oxford: Blackwell.
Ho, M. L., & Platt, J. (1993). *Dynamics of a contact continuum: Singaporean English*. Oxford: Oxford University Press.
Hudson, T. (1993). Nothing does not equal zero: Problems with applying developmental sequence findings to assessment and pedagogy. *Studies in Second Language Acquisition*, *15*, 461–493.
Huebner, T. (1983). *A longitudinal analysis of the acquisition of English*. Ann Arbor, MI: Karoma.
Ionin, T., Ko, H., & Wexler, K. (2003). Specificity as a grammatical notion: Evidence from L2 English article use. In G. Garding and M. Tsujimura (Eds.). *WCCFL 22 Proceedings* (pp. 245–258). Somerville, MA: Cascadilla Press.
Jarvis, S. (2002). Topic continuity in L2 English article use. *Studies in Second Language Acquisition*, *24*, 387–418.
Kachru, B. (1965). The indianness of Indian English. *Word*, *21*, 391–410.
Kachru, B. (1983). *The indianization of English*: *The English language in India*. Oxford: Oxford University Press.
Kachru, Y. (2003). On definite reference in world Englishes. *World Englishes*, *22*, 497–510.
Kandiah, T. (1991). South Asia. In J. Cheshire (Ed.), *English around the world* (pp. 271–288). New York: Cambridge University Press.
Klein, F. (1980). A quantitative study of syntactic and pragmatic indications of change in the Spanish of bilinguals in the U.S. In W. Labov (Ed.), *Locating language in time and space* (pp. 69–82). San Diego, CA: Academic Press.
Klein, W., & Perdue, C. (1992). *Utterance structure*: *Developing grammars again*. Amsterdam: Benjamins.
Lambrecht, K. (1994). *Information structure and sentence form*. New York: Cambridge University Press.
Lowenberg, P. (1986). Nonnative varieties of English: Nativization, norms, and implications. *Studies in Second Language Acquisition*, *8*, 1–18.
Lyons, C. (1999). *Definiteness*. New York: Cambridge University Press.
McGregor, R. S. (1995). *Outline of Hindi grammar with exercises*. Oxford: Oxford University Press.
Mesthrie, R. (1992). *English in language shift*. New York: Cambridge University Press.
Mishra, A. (1982). Discovering connections. In J. Gumperz (Ed.), *Language and social identity* (pp. 57–72). New York: Cambridge University Press.
Mougeon, R., & Beniak, E. (1996). Social class and language variation in bilingual speech communities. In D. Schiffrin, G. Guy, J. Baugh, & C. Feagin (Eds.), *Towards a social science of language*: *Vol. 1. A festschrift for William Labov* (pp. 69–99). Amsterdam: Benjamins.
Parrish, B. (1987). A new look at methodologies in the study of article acquisition for learners of ESL. *Language Learning*, *37*, 361–383.
Pienemann, M., Johnston, M., & Meisel, J. M. (1993). The multidimensional model, linguistic profiling, and related issues. *Studies in Second Language Acquisition*, *15*, 495–503.
Platt, J., Weber, H., & Ho, M. L. (1984). *The new Englishes*. London: Routledge.
Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255). San Diego, CA: Academic Press.
Rickford, J. (2002). Implicational scales. In J. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *Handbook of language variation and change* (pp. 142–167). Oxford: Blackwell.
Rousseau, P., & Sankoff, D. (1978). Advances in variable rule methodology. In D. Sankoff (Ed.), *Linguistic variation: Models and methods* (pp. 57–69). San Diego, CA: Academic Press.
Sahgal, A. (1991). Patterns of language use in a bilingual setting in India. In J. Cheshire (Ed.), *English around the world* (pp. 299–308). New York: Cambridge University Press.
Sankoff, G. (1983). Comments on Valdman's "Creolization and second language acquisition." In R. Andersen (Ed.), *Pidginization and creolization as second language acquisition* (pp. 235–240). Rowley, MA: Newbury House.
Sankoff, G., & Mazzie, C. (1991). Determining noun phrases in Tok Pisin. *Journal of Pidgin and Creole Languages*, *6*, 1–24.
Schiffman, H. (1999). *A reference grammar of spoken Tamil*. New York: Cambridge University Press.
Sharma, D. (2005). Dialect stabilization and speaker awareness in nonnative varieties of English. *Journal of Sociolinguistics*, *9*, 194–224.
Sridhar, K. K. (1985). Sociolinguistics of nonnative English. *Lingua*, *68*, 39–58.

Sridhar, S. N. (1990). *Kannada*. London: Routledge.
Tarone, E., & Parrish, B. (1988). Task-related variation in interlanguage: The case of articles. *Language Learning*, *38*, 21–45.
Trenkic, D. (2002). Establishing the definiteness status of referents in dialogue (in languages with and without articles). *Working Papers in English and Applied Linguistics 7* (pp. 107–131). Cambridge: Research Centre for English and Applied Linguistics, University of Cambridge.
Vallduví, E. (1992). *The informational component*. New York: Garland.
Williams, J. (1987). Nonnative varieties of English: A special case of language acquisition. *English World-Wide*, *8*, 161–199.
Wolfram, W. (1985). Variability in tense marking: A case for the obvious. *Language Learning*, *35*, 229–253.
Young, R. (1996), Form-function relations in articles in English interlanguage. In R. Bayley & D. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 135–175). Amsterdam: Benjamins.