Studies in Second Language Acquisition, 2015, **37**, 101–133.
doi:10.1017/S0272263114000357

EMBEDDED *WH*-QUESTIONS IN L2 ENGLISH IN INDIA

Inversion as a main clause phenomenon

David Stringer
Indiana University

This corpus study brings a second language (L2) research perspective, insights from generative grammar, and new empirical evidence to bear on a long-accepted claim in the World Englishes literature—namely, that inversion with *wh*-movement in colloquial Indian English is obligatory in embedded clauses and impossible in main clauses. It is argued that this register of Indian English is a L2 variety, functioning as part of a multilingual code repertoire, but that syntactic universals apply to first and second languages alike. Despite recent attempts at formalization, this distribution should be unattested, as such a grammar would fall outside the constraints of Universal Grammar and would contradict proposed discourse-pragmatic principles of natural language. A Perl program was created to search the Indian subcorpus of the International Corpus of English (Greenbaum, 1996) for relevant distributional patterns. Results reveal that *wh*-inversion in Indian English operates in the same way as in other varieties: It is robustly attested in main clauses and appears only occasionally in embedded clauses where syntactic and pragmatic conditions allow; it is obligatory only with interrogative complementizer

Special thanks are owed to three people at Indiana University, Bloomington for their assistance with various parts of this project: Markus Dickinson, for writing the original Perl program and teaching me how to tweak it; Stephanie Dickinson, for her invaluable help with statistics; and Yi-Ting Wang, for her excellent support in the initial stage of manual filtering. Thanks are also due to those who have given feedback on these ideas: Kathleen Bardovi-Harlig, Joe Emonds, Phil Lesourd, and especially Lucia Pozzan, for sharing her own insightful work on this topic.

Correspondence concerning this article should be addressed to David Stringer, Department of Second Language Studies, Memorial Hall 310, Indiana University, 1021 E. Third Street, Bloomington, IN 47405. E-mail: ds6@indiana.edu

deletion. Thus, contrary to the standard account but commensurate with recent corpus studies, users of English in India exhibit knowledge of universal constraints in this domain.

One purpose of this article is to report an investigation into the syntax of English as a second language (L2) in India from the perspective of generative approaches to acquisition. A second objective is to establish more generally that the field of SLA stands to gain by broadening its purview to include users of English in postcolonial environments.[1] The empirical part of the study involves a topic with a long history in generative research—namely, main versus embedded clause phenomena (Emonds, 1970, 1976; Hooper & Thompson, 1973; Ross, 1973), which has recently been the focus of renewed interest (Aelbrecht, Haegeman, & Nye, 2012; Haegeman, 2012). The question is whether it is possible that L2 varieties of English violate the universal constraint by which no grammatical process may be found only in embedded clauses. The corpus study reported here challenges a conventional assumption in the World Englishes literature concerning the grammar of L2 English in India: that auxiliary inversion with *wh*-movement (henceforth *wh*-inversion) is exactly the opposite of standard U.S. and U.K. English, being impossible in main clauses yet obligatory in embedded clauses (Bhatt, 2000; Mesthrie & Bhatt, 2008; Trudgill & Hannah, 2008).[2] If the prevailing account were valid, this language variety would constitute a rogue grammar in the sense of Thomas (1991). However, the results suggest that, in this respect, the grammars of L2 users of English in multilingual societies, just like those of monolingual speakers of standard varieties, fall within the compass of Universal Grammar (UG).

The second aspect of this study is, in some sense, an obvious extension of previous generative L2 research. If all L2 varieties are constrained by UG (Dekydtspotter, 2009; Montrul, 2004; Schwartz & Sprouse, 1996; Slabakova, 2008; White, 2003), and if studying the interplay of languages in the bilingual mind can shed light on the nature of the language faculty, then much can be learned from moving beyond the (often) more readily available student populations in countries such as the United States and the United Kingdom into the wider world of learners and users of L2 English in postcolonial societies in Asia and Africa, which represent, by far, the greatest number of regular users of English as an additional language. At least 400 million people (Crystal, 2006) acquire English in late childhood and adulthood in postcolonial societies such as Ghana, India, Kenya, Malaysia, Nigeria, Pakistan, the Philippines, and South Africa, primarily to communicate intranationally rather than to approximate to first language (L1) norms, which calls for a more pluricentric perspective on the processes and goals of language acquisition

than is found in mainstream SLA.[3] This enterprise also involves engaging with the World Englishes literature and reevaluating traditional SLA notions such as *interlanguage*, *target language*, and *native speaker*, in contexts in which the target is not approximation to monolingual norms, as in many Western countries, but rather acquisition of a multilingual code repertoire.

L2 ENGLISH IN INDIA: INITIAL PREMISES

Any formal syntactic investigation of language in the populations that have almost exclusively been studied from the sociolinguistic perspective of the World Englishes tradition risks misinterpretation, as the same variety may be considered a L1 or a L2, a learner language or a user language, an indigenized dialect or an emerging standard, depending on the nature of the investigation and the particular research tradition. From a generative L2 perspective, sociopolitical designations are not the primary concern. All natural languages are expressions of a universal human language faculty, and no variety is considered comprehensively superior to any other. Before proceeding further, certain terminology will be made clear in the context of the current approach.

First, the terms *L1* and *L2* will be maintained with their intended interpretation in the mainstream acquisition literature, in which they are generally defined not in terms of relative language dominance but in terms of the age of the language learner. In a comprehensive review of the child L2 literature, Unsworth (2005) synthesized available studies to define L1 acquisition as occurring between 0 and 4 years old, and the special case of child L2 acquisition as occurring between 4 and 7 years old. Proponents of a critical period for syntax assume that acquisition becomes compromised at around 4 to 7 years old: Johnson and Newport (1989) suggested 7 years old; Long (2003) proposed 6 years old; and Meisel (2009, p. 10) specified it as "during the last months of the fourth year and perhaps shortly afterwards," whereas Newport and Supalla (1990) settled for a period between 4 and 6 years old. In this study, I follow Schwartz and Sprouse (1996), White (2003), and others in assuming that, although L2 acquisition is complicated by crosslinguistic influence, there is, in fact, no biologically determined critical period for syntax.

The percentage of the population in India for whom English is a language in the home in the preschool years is almost negligible and has remained fairly constant for the last 40 years, as revealed in official census data (1971: 0.03%; 1981: 0.04%; 1991: 0.02%; 2001: 0.02%).[4] By standard definitions, therefore, almost all speakers of English in India are L2 users of the language, being exposed to it through the school system, with the syntax of their L1(s) already in place. It should be stressed that this distinction has nothing to say about eventual proficiency,

as, for a minority of speakers, the L2 may become the dominant language in adulthood, and for many it will serve as an integral component of a sociofunctional code repertoire (B. B. Kachru, 1986/1990, pp. 57–80). It should also be noted that defining this variety as a L2 is independent of the debate as to whether the target for L2 acquisition is a standard established overseas (e.g., U.S. or U.K. English) or an indigenized, norm-developing variety. Second language acquisition as a field has yet to take seriously the fact that a L2, taught by L2 users to L2 learners, without significant L1 input, can be elevated to the status of national language (although note of this was made by Selinker, 1972, p. 217). Indeed, the population of L2 English users in India is staggering in size and sociopolitical influence. The 2001 census furnished a 700-page supplement on multilingualism that found that 11% of the population (almost 120 million people) reported being conversant in English as a second or third language. This ignores the fact that, for many, English is a fourth or fifth language, and it is widely believed that the self-reports of the census considerably underestimate the number of speakers, the true figure being about double the official estimate (B. B. Kachru, 2006, p. 467, fn. 10).

A second set of terms that require initial clarification includes *interlanguage*, *target language*, and *native speaker*. *Interlanguage* (Selinker, 1972) remains a widespread and accepted term in SLA denoting the L2 system; it is a term that for many years has had strictly positive connotations in the field. It is used to convey the idea that L2s are most usefully studied not as impoverished versions of target languages but as natural grammars in their own right. As stated by Adjemian (1976), "underlying the IL [interlanguage] hypothesis is the unwritten assumption that ILs are linguistic systems in the same way that natural languages are. . . . ILs are natural languages" (p. 298; see also Eckman, 1981; Gass, 2013; Schwartz & Sprouse, 1996; Sharwood Smith, 1988; White, 2003). This perspective ought to be welcomed by researchers who value the linguistic knowledge of speakers of multiple languages independently of externally imposed prescriptive standards. However, it should be noted that in the World Englishes tradition, the term *interlanguage* lacks positive connotations, and Selinker's (1972) early characterization of Indian English interlanguage in terms of "fossilized competence" (p. 217) continues to cause umbrage. Sridhar and Sridhar (1986) and Mesthrie and Bhatt (2008) consider interlanguage to be an inappropriate concept with negative connotations; it is seen as neither one thing nor the other—an approximative system made up of bits and pieces of L1 and L2 that is variable and unstable, particular to an individual mind rather than a group phenomenon, and unsuited for more stable L2 varieties such as Indian English. In the current article, *interlanguage* is to be understood in the standard SLA sense of the term. In other words, I assume that interlanguages are indeed natural languages in the same sense as L1s, that they are systematic, and that language universals

ought to be just as relevant to multilingual code repertoires as they are to monolingual language systems.

The terms *target language* and *native speaker*, however, are not directly applicable to the study of the linguistic knowledge of L2 users of postcolonial Englishes. In most postcolonial contexts, the target is functional multilingualism rather than a single language. Although knowledge of two or three languages is perhaps the norm, some communities in Mumbai may use as many as five or six languages just to get through the week, with the vocabulary, syntax, and available registers for each one tailored to the needs of the contexts of use (B. B. Kachru, 1986/1990, Chapter 4). For children and adults acquiring L2s outside the home, the target is often a code repertoire, rather than the kind of one-language-fits-all-functions model of the monolingual native speaker (for further discussion from the World Englishes perspective, see B. B. Kachru, 1985; Y. Kachru, 1994; Singh, 1998; for critiques of the concept of the native speaker within SLA theory, see Cook 1991, 2002, 2003; Ortega, 2005, 2013).

In any case, for the purposes of research into language universals, the lack of an idealized single target language or an idealized native-speaker model is irrelevant. All interlanguage systems in a code repertoire, by hypothesis, are constrained by UG and involve universal parsing mechanisms. Thus, a wealth of data from L2 users of English in postcolonial societies with a vast range of knowledge levels and a rich variety of language combinations can potentially be brought to bear on current issues of L2 acquisition and knowledge.

In the following section, attention turns to the question under investigation in the current study: whether supposedly universal patterns of embedding and inversion hold in L2 English in India or whether this variety breaks the mold such that late acquisition results in a system not subject to the same constraints as monolingual systems.

WH-INVERSION IN INDIAN ENGLISH AS A MIRROR OF U.K. AND U.S. ENGLISH

A Formalization of the Proposed Distribution

The long-standing claim that this study attempts to challenge is that colloquial Indian English syntax does not allow inversion in main clause questions but has obligatory inversion in embedded questions, as illustrated in examples (1) and (2) (Trudgill & Hannah, 2008, p. 137, Examples 9a, 9b):

(1) a. *What this is made from?*
    b. *Who you have come to see?*
(2) a. *I asked him where does he work.*
    b. *I wonder where is he.*

Such forms have been noted since the early days of scholarship on Indian English (Whitworth, 1907/2007, pp. 148–150) and are regularly supplied in descriptive grammatical overviews of the language. Trudgill and Hannah (2008) characterize this aspect of syntax as "exactly the opposite of EngEng [U.K. English] usage" (p. 137), and Mesthrie and Bhatt (2008) state that "the generalization of question-formation strategy in colloquial Indian English is the mirror opposite of the generalization of question-formation strategy in Std Ind [Standard Indian] English" (p. 98). That such forms exist is without question. As shall be shown, they exist in several varieties of English. The issue is whether the alternative formulations are ruled out, such that, as Bhatt (2004/2008) states, "the simple empirical generalization that emerges from the data . . . is that in Vernacular IndEng *inversion is restricted to embedded questions* [emphasis added]; it does not apply in matrix questions" (p. 551).

Bhatt (2000) and Mesthrie and Bhatt (2008, pp. 96–106) go beyond anecdotal observation in providing a formal account of this variation, contrasting Standard Indian English with Colloquial Indian English.[5] They elaborate an analysis in terms of optimality theory (OT; Grimshaw, 1997; Prince & Smolensky, 1993) to distinguish two distinct formal grammars: one with inversion restricted to and obligatory in main clauses and the other with inversion restricted to and obligatory in embedded clauses. As a complete evaluation of this account is beyond the scope of this article, the following description conveys the basics of the analysis.

The OT framework conceives of grammars as hierarchies of constraints on well-formedness. Particular language grammars differ from one another in how they rank the constraints. These constraints are not inviolable; rather, a constraint may be violated in the case that the structure is preferred according to the values of a higher ranking constraint. In any language, if two syntactic variations are considered, evaluation of the constraints violated in terms of their ranking determines which option is considered "optimal." In this case, the relevant constraints are as follows:

1. OP-SPEC: Operators must be in specifier (Spec) position.
2. STAY: Constituents must remain in position; no movement is allowed.
3. OB-HD: Heads of selected projections must be filled, either overtly or by a trace (*t*).

In Bhatt's (2000) and Mesthrie and Bhatt's (2008) analysis of main clauses, the first two constraints are relevant. The ranking for Standard English(es) is OP-SPEC » STAY, so a sentence such as *What would you like to eat?* is preferred over *What you would like to eat?* In the sentence *What would you like to eat?* OP-SPEC is respected, due to *wh*-movement to the specifier of the complementizer phrase (Spec, CP), even though

STAY is violated twice: once by movement from T (tense) to C (complementizer) and again by movement from object position to Spec, CP. However, in *What you would like to eat?* which is assumed to be generated by adjunction of the *wh*-phrase to the tense phrase (TP), OP-SPEC is violated, ruling out the candidate structure. In contrast, in Indian English, the proposed ranking is STAY » OP-SPEC, such that the preferred candidate is *What you would like to eat?* In this case, although both STAY and OP-SPEC are each violated once, the other candidate is worse: two cases of movement (the modal and the *wh*-phrase) would amount to two violations of the higher ranking constraint.

In the proposed analysis of embedded clauses, only the last two constraints are relevant, as the *wh*-word is assumed to have moved to Spec, CP in both language varieties. The ranking for Standard English(es) is STAY » OB-HD, so a sentence such as *I wonder what he is eating*, with only one instance of movement, is preferred over *I wonder what is he eating*, which involves two violations of STAY. That OB-HD is violated by *I wonder what he is eating*, as the head of the CP remains unfilled, does not matter, because STAY is the higher ranking constraint. In Indian English, however, the ranking is argued to be OB-HD » STAY, which means that *I wonder what is he eating* emerges as the stronger candidate. In this case, OB-HD is not violated, as T-to-C movement results in the head of the CP being overtly filled.

For a more complete explanation of this account including OT tableaux, see Mesthrie and Bhatt (2008, pp. 96–106). For current purposes, it is sufficient to note that the logic of argumentation is sound given the assumptions of the analysis, both overt (the OT constraints and rankings and the adjunction analysis of *wh*-fronting) and covert (that variation is due to two separate grammars and that there is no optionality involved). Moreover, the general spirit of inquiry, to demonstrate that "Indian vernacular English is just as systematic and logical as any other variety of English" (Bhatt, 2000, p. 69), is one that I endorse unconditionally.

However, two problems arise that suggest that an alternative analysis is needed. First, despite commonly repeated claims in textbooks and anecdotal reports, there is a lack of empirical support for this distribution. Second, a language with *wh*-inversion entirely restricted to embedded clauses would stand as an anomaly, given previous research on universal patterns of embedding and inversion.

The Status of Existing Evidence

To date, no systematic empirical study in the World Englishes paradigm has shown this distribution to exist, despite various anecdotal claims

and informal investigations. Bhatt's (2000) evaluation of inversion in embedded clauses was based on presenting only six tokens of this construction, as part of a series of 24 sentences all exemplifying Indian English variants, to 27 people (including high school English teachers, two doctors, two linguists, and an engineer) and asking them if the sentences were likely to be produced by a speaker of Indian English or British English.[6] The results for inversion in embedded clauses were as follows: 81% were judged to be Indian English, 0% British English, and 19% neither (i.e., participants could not imagine any inversion in embedded clauses). However, interpreting these results is less than straightforward, for a number of reasons: (a) the numbers of participants and stimuli were very small; (b) the test instrument was not balanced; (c) participants gave not grammaticality judgments but rather perceptions of what sounded "more Indian" and "less British;" and (d) the fact that these forms are attested in Indian English in no way implies that noninversion in embedded sentences is disallowed.

The interpretation of obligatory inversion in embedded clauses also flies in the face of recent corpus studies that suggest that noninversion in embedded clauses is the preferred (though not the only) option in Indian English. Although these investigations may be considered problematic in various respects, they do all point to the same conclusion. In a pioneering corpus linguistic investigation of this issue, Hilbert (2008) performed a custom search using the International Corpus of English, henceforth ICE (for discussion of this set of freely available online corpora, see Greenbaum, 1996). She searched the spoken files of the Indian subcorpus (ICE-India), which contain 600,000 words, for patterns of inversion in main and embedded clauses. The rates of *wh*-inversion reported for main clauses were 81% inversion (540 tokens) and 11% noninversion (76 tokens); the proportions for embedded clauses were 18% inversion and 82% noninversion. However, the precise calculation of the figures remains unclear. There were no stated criteria for exclusion, thus no account was given of the apparent exclusions in main clauses ($11 + 81 < 100\%$), and no raw numbers were given for embedded clauses. The importance of this will become apparent in the Results section; in the current investigation, as many as 35% of strings of *wh*-elements followed by auxiliaries turned out not to be *wh*-questions. Possible hits that are candidates for exclusion include *wh*-clefts, relative clauses, and cases that are ambiguous with respect to movement, such as subject *wh*-questions and copied auxiliaries or modals (e.g., *How should we should count these?*). Another issue in Hilbert (2008) was that *yes/no* questions were identified in the untagged corpus by searching for transcribed question marks. However, not all *yes/no* questions in the corpus are followed by a question mark (many are inside longer stretches of dialogue). The only information concerning the search for *wh*-questions is that they "were selected independently of this mark-up" (Hilbert, 2008, p. 270).

In a subsequent study, drawing on a subcorpus of the spoken files of ICE-India—namely, the "private conversations" (S1A-001 to S1A-100)—Hilbert (2011) focused on how inversion may be conditioned by the verb type: In main clauses, inversion is preferred, but more so with *be* (about 95%) than with modals or auxiliaries (about 80%) or with lexical verbs involving *do*-support (about 60%). Neither raw numbers nor percentages were provided; the preceding figures are estimations based on a bar graph (Hilbert, 2011, p. 130, Figure 2). Hilbert argued that, in embedded *wh*-questions, inversion is almost entirely restricted to third-person singular *is*, leading her to propose an analysis in terms of fixed chunks derived from a process of learning by imitation (to be examined in more detail in the Discussion section). However, these findings are difficult to interpret, as there are no raw numbers, no accurate percentages, no statistics, and, again, no information concerning the search criteria.

Balasubramanian (2009) also found that the proportion of noninversion with *wh*-phrases in main clauses in the spoken files of ICE-India was very low, at only 8% (72/880), although a more detailed breakdown is lacking, and the analysis did not extend to embedded clauses. In this case, the search followed the tagging of the corpus. However, as in Hilbert (2008), there was no reference to any criteria for filtering.

Sedlatschek (2009) reported findings from a smaller corpus of 180,000 words culled from newspapers, broadcasts, and student essays. In this case, the author's methodology was more explicit (p. 292). First, all *wh*-clauses in the corpus were identified using the WordSmith concordance program (Scott, 2004), with nine different *wh*-elements as keywords. The data were then filtered manually, eliminating relative clauses, subject *wh*-questions, and instances of usage that conform to Standard English conventions (such as noninversion following *how come*). The remaining data were categorized into main and embedded clauses. Proportions of embedded *wh*-inversion in embedded clauses from the three subsections of the corpus were as follows: 10% from the press (2/20), 13.3% from broadcasts (10/75), and 2.5% from essays (1/40) (p. 297). Despite the limitations of scale and a reliance on written rather than oral production, the analysis was careful, the distribution was clear, and the findings were in accordance with the other corpus studies cited.

Investigations such as Balasubramanian (2009), Hilbert (2008), and Sedlatschek (2009) stand as important pioneering studies and deal with more issues than those addressed within the narrow focus of the current investigation. They highlight an exciting new avenue in corpus-based research, which promises not only to shed greater light on the nature of the so-called New Englishes but also to facilitate comparison with classroom varieties of English (see Mukherjee & Hundt, 2011). The preceding review makes clear that, in addition to a general lack of statistical analysis, each of these initial findings requires further investigation due to partial reporting of the results, because the search methodology was unclear,

or because there were no clear criteria for inclusion or exclusion. However, they all trend in the same direction. On the issue of *wh*-inversion in Indian English, the current study hopes to provide more comprehensive empirical evidence and to elucidate the relevance of the results for L2 research on language universals.

UNIVERSAL PATTERNS OF EMBEDDING AND INVERSION

Syntactic and Discourse-Pragmatic Constraints

Whether Indian English exhibits *wh*-inversion in embedded clauses is not only an empirical question about a particular language variety; it also bears on broader issues of syntactic universals and discourse-pragmatic constraints. The hypothesized syntactic distribution is implausible whether viewed from a formalist perspective (given the history of research on main clause phenomena) or a functionalist perspective (considering the backgrounding of information in embedded clauses).

Several decades of formal research on main clause phenomena lead us to expect that no natural language should allow inversion in embedded clauses while ruling it out in main clauses. The first systematic investigation of the syntax of main and embedded clauses was that of Emonds (1970, 1976). Emonds (1970) proposed the structure-preserving constraint, an important feature of which was the restriction of certain syntactic phenomena, termed *root transformations*, to main clauses. On this approach, *wh*-inversion is a root transformation, and any apparent exceptions must be otherwise explained. Emonds (1976, pp. 23–25) considered apparent counterexamples such as *wh*-inversion in embedded clauses in African-American English, as shown in (3) (Labov, Cohen, Robbens, & Lewis, 1968, p. 299, Examples 415, 419).

(3) a. *I don't know how did I do it.*
b. *Where did she get the coat from I don't know.*

He suggested that such forms may be similar in structure to the "mixed indirect discourse" (p. 25) of German, in which the apparent embedded clause exhibits verb-second (V2) syntax, which is ordinarily a main clause phenomenon (in today's terminology, the finite verb raises to C, and any single argument or adjunct phrase may raise to Spec, CP). On this account, all apparent cases of *wh*-inversion in embedded clauses are actually derived parenthetically from the fusion of two main clauses, as in the following example (Emonds, 1976, p. 25, Examples 11, 12).

(4) a. *John wondered (mused), (why) should he be early.*
b. *John wondered thus: Why should he be early.*

Subsequent to Emonds's (1970) proposal, Ross (1973) developed an alternative statement of constraints on embedding. Referring to main clauses as "upstairs" and embedded clauses as "downstairs," a jargon still familiar in generative linguistics today, he proposed "a principle whose truth is borne out in myriad cases of Real Apartment Life, and which is therefore dubbed: (1) *The Penthouse Principle*: More goes on upstairs than downstairs" (p. 397). He proceeded to recast it more soberly as follows:

(5) No syntactic process can apply only in subordinate clauses.

Ross (1973) argued that apparent exceptions such as NP-raising (noun phrase raising) and *that*-deletion crucially involve activation of another, higher clause, and syntactic processes such as reference questions, fronting rules, embedded force rules, pseudoimperatives, and various idioms either do not apply in embedded clauses or "are more active, have a wider scope of application . . . when they operate upstairs" (p. 401). Thus, although (5) remains valid, the final statement of the penthouse principle is, in implicational terms: "where ≥ means 'is as syntactically active as, or more syntactically active than,' the following inequality obtains: Main clauses ≥ subordinate clauses" (p. 411). The penthouse principle makes different predictions from the structure-preserving constraint. Whereas the structure-preserving constraint implies that *wh*-inversion should never be found in embedded clauses, the penthouse principle allows for this phenomenon on the condition that it applies freely in main clauses and that its use in embedded clauses is relatively restricted.

In an early functionalist response to these syntactic accounts of the asymmetry between main and embedded clauses, Hooper and Thompson (1973) argued that discourse factors were more useful in explaining such differences. Embedded clauses generally contain backgrounded information, which is less likely to undergo the kind of transformations associated with topicalization, contrast, and presentative focus. However, on this account, such transformations remain possible when the embedded information involves the assertion of a proposition, as shown in examples (6)–(8) (Hooper & Thompson, 1973, p. 466, Example 5; p. 478, Example 92; p. 480, Example 123).

(6) *Alice vowed that under no circumstances would she loan me the key.*
(7) *It appears that this book he read thoroughly.*
(8) *The scout discovered that beyond the next hill stood a large fortress.*

As for questions, actual requests for information are fronted in discourse and appear in main clauses, but there are circumstances in which, depending on the type of predicate and the discourse context, the embedded interrogative is also treated as a "real question." G. M. Green (1996) argued

that pragmatics sheds light on precisely which contexts are felicitous for embedded inversion in colloquial American English. She observed that embedded questions such as (9) and the truncation of embedded questions ("sluicing") as in (10) both imply that "the individual to whom the answer is implied or assumed to be relevant is in fact ignorant of the answer" (p. 141). Thus the (a) sentences are better than the (b) sentences in (9) and (10) (G. M. Green, 1996, p. 141, Examples 23a, 23b, 24a, 24b).

(9) a. *She wants to know who did I appoint.*
b. ?*She already knows who did I appoint.*
(10) a. *John broke something but he won't say what.*
b. ?*John broke something and he said what.*

In a more recent discussion of main and embedded clauses, Bybee (2002)[7] explained the thrust of this discourse-pragmatic approach to the difference between clauses types as follows: "Typical subordinate clauses are pragmatically flat just as they tend to be intonationally flat and less susceptible to permutations for pragmatic purposes" (p. 5). Unlike main clauses, which provide the default location for topicalization, contrast, focus, and questions, embedded clauses have functions that are generally "much more modest" (p. 4): modifying or further identifying the head NP in the case of relative clauses or modifying the main event in terms of causes, times, or locations in the case of adverbial clauses.

The implications of this approach for syntactic distribution seem to align with the penthouse principle, which makes predictions in terms of proportional distribution, rather than the categorical restriction implied by the structure-preserving constraint. Both Hooper and Thompson (1973) and Bybee (2002) argue that an approach based on discourse pragmatics is more explanatory than essentially descriptive syntactic accounts pairing transformations with grammatical domains, which are too constrained and cannot convincingly account for the counterexamples. In contrast, proponents of a syntactic approach tend to see purely pragmatic accounts that rely on "vague discourse properties" (Emonds, 2012, p. 28), such as assertion and presentational focus, as failing to account for syntactic constraints in embedded clauses.

A possible bridge between formal syntactic and discourse considerations has emerged in work by Emonds (2004, 2012), who argues that the previous category of sentential *root* may be reconstrued as a discourse projection, defined as an unselected finite inflection phrase (IP)—that is, not an argument or adjunct of an underlying lexical item. The term *discourse projection* captures the fact that "the constituents of a root clause stand in privileged relations with surrounding categories of discourse" (Emonds, 2004, p. 81) such as focus, topic, and speaker-oriented interjections. Various types of movement to the left edge of the main clause are analyzed in terms of category-less phrases (XPs) characterized as

discourse shells, which dominate discourse projections and whose Spec positions function as unrestricted landing sites for root phenomena. The concept of discourse shells is intended to generalize over and replace previous proposals of topic phrases and focus phrases (Rizzi, 1997, 2006).[8]

Regarding seemingly exceptional transformations in embedded clauses, Emonds (2004, 2012) argues that many are, in fact, more restricted than generally realized. Cases of what he terms RIDE (root-like indirect discourse embedding) are always finite; they substitute for complements, not adjuncts; they are usually introduced by verbal or adjectival heads, not nominal or prepositional heads; and one argument of the introductory predicate must be animate. The distinction between root and nonroot phenomena is maintained but is loosened somewhat, in that individual languages may allow certain types of embedded clauses as discourse projections as long as RIDE conditions are met and with a degree of parametric variation as to the types of clauses involved. Advantages of this approach over a pure discourse-pragmatic account include a formal characterization of constraints on RIDE and a coherent structural analysis of the phenomena.

There is another possible explanation of certain main versus embedded clause discrepancies, which, although very simple, appears to account for a significant amount of the variation found with *wh*-inversion. Both Emonds (2004, 2012) and the cartographic approaches based on Rizzi (1997, 2006) eschew a simple, unified CP analysis of question formation in main and embedded clauses. Yet there appears to be a straightforward link between the existence of overt complementizers and cases of prohibition on movement in embedded clauses—as shall be shown later in this article—which is easily explained if the structure used for complement clauses and that used as a landing site in question formation is one and the same functional projection. As Lightfoot (2012) puts it, "the reason [for restrictions on syntactic operations in embedded clauses] is that the syntactic enabling conditions are not met" (p. 172). The results of the current study support this conclusion.

To summarize: Across a broad range of linguistic frameworks, whether one adopts an account in terms of general syntactic principles or in terms of discourse-pragmatic constraints, there is consensus in the literature as to the well-documented asymmetry between main and embedded clauses, and as such the traditional claim about inversion in Indian English stands out as an anomaly.

Embedded Inversion in L1 Dialects

Evidence from L1 dialects and colloquial forms of standard varieties can also shed light on the nature of embedded inversion. In various L1

dialects, embedded inversion is relatively commonplace, but only when specific syntactic and pragmatic conditions obtain, and always in conjunction with (at least optional) inversion in main clauses. Belfast English has true embedded inversion, as argued by Henry (1995, pp. 105–123). Example (11) shows the sequence of tense and change of pronouns, and (12) shows inversion in interrogative complements of verbs that cannot select direct speech (Henry, 1995, pp. 106–107, Examples 11a, 11b, 19, 20).

(11) a. *They asked me, "Have you read* War and Peace*?"*
b. *They asked me had I read* War and Peace*.*
(12) a. *The police found out had the goods been stolen.*
b. *We couldn't establish did he meet them.*

However, inversion is not a feature of all embedded questions. If the clause is introduced by an interrogative complementizer,[9] inversion is impossible, as shown in (13), and although *wh*-movement may co-occur with the complementizer *that* in this variety, such combinations rule out inversion, as shown in (14) (adapted from Henry, 1995, pp. 107–108, Examples 25a, 25b, 25c, 27, 33).

(13) a. *They couldn't work out {whether/if} we had left.*
b. *They couldn't work out had we left.*
c. **They couldn't work out {whether/if} had we left.*
(14) a. *I wonder which dish that they picked.*
b. *I wonder which dish did they pick.*
c. **I wonder which dish that did they pick.*

These examples all point to inversion by means of a CP at the head of an embedded clause, rather than the fusion of two main clauses. A further possibility in Belfast English that cements the embedded clause analysis involves cases with long-distance movement, as shown in (15), which cannot be analyzed as parentheticals (adapted from Henry, 1995, pp. 108–109, Examples 35, 46).

(15) a. *Who$_{\mathbf{i}}$ did John hope would he see* $\boldsymbol{t}_{\mathbf{i}}$?
b. *Who$_{\mathbf{i}}$ did John say did Mary claim had John feared would Bill attack* $\boldsymbol{t}_{\mathbf{i}}$?

McCloskey (2006, pp. 89–90) gives the following examples from Irish English (Hiberno English), which he notes are not parentheticals. Note the pronominal binding in (16) and the impossibility of declaratives with a rising intonation pattern in this context in (17a), in contrast with their acceptability in true parentheticals, such as in (17b) (McCloskey, 2006, p. 89, Example 8; p. 90, Examples 13–14).[10]

(16) *Every male physicist wonders will he be awarded a Nobel Prize.*

(17) a. **I wonder it's raining?*
b. *It's raining?, she mused.*

McCloskey (2006) goes on to observe that, in this variety, embedded inversion is usually found with question predicates such as *ask* / *wonder* / *inquire* / *want to know* / *want to see*, but it is characteristic of resolutive predicates such as *find out* / *know* / *remember* / *discover* only when they are interrogative or negative (pp. 112–115). Moreover, McCloskey notes that embedded inversion is found more generally in the presence of "devices which determine nonveridical contexts" (p. 114). The same principle appears to apply in African-American English, as previously shown in the example in (3) for negation and in the examples in (18) with imperatives and desideratives (L. J. Green, 2002, pp. 87–88, Examples 20f, 20h, 20e, 20k).

(18) a. *Tell me do it make any sense.*
b. *Go over there and see did they bring my car in.*
c. *I wanted to see was it the one we bought.*
d. *I wanted to know could they do it for me.*

That African-American English also allows noninversion in main clauses does not run counter to the penthouse principle, as such inversion is optional. No speaker disallows inversion in main clauses; in fact, matrix inversion is one of three typical patterns (L. J. Green, 2002, p. 87).

A further observation may be made with examples from Tyneside English (my own childhood dialect), spoken in the northeast of England. First, not only is inversion possible with interrogative complementizer deletion, but it is obligatory, this being the only manifestation of the embedded clause's interrogative force, as illustrated in (19). For resolutive verbs, which, unlike question verbs, select both *that* and *if*/*whether*, there is, in fact, no ambiguity in embedded clauses: Inversion indicates *if*-deletion, and noninversion indicates *that*-deletion, as in (20).

(19) a. *Terry asked Julie would she like to gan to the beach.*[11]
b. **Terry asked Julie she would like to gan to the beach.*
(20) a. *Me mam wants to know was me dad happy yesterday.* (*if*-deletion)
b. *Me mam wants to know me dad was happy yesterday.* (*that*-deletion)

Such distribution appears to hold in the other dialects considered here (L. J. Green, personal communication, August 6, 2010; A. Henry, personal communication, June 13, 2011; J. McCloskey, personal communication, June 23, 2011). The relevant syntactic condition appears to be as follows:

(21) In languages with inversion in matrix clauses, optional deletion of an interrogative complementizer results in obligatory inversion in embedded clauses.

This brief survey of L1 dialects has established that inversion in embedded clauses is a relatively widespread phenomenon that is not particular to Indian English and is everywhere subject to certain syntactic and pragmatic constraints, as would be expected on any formal analysis in which the CP projection is the target of the relevant movement operations. None of the varieties discussed violate the penthouse principle, in that inversion is allowed freely in main clauses, even though it is not always obligatory.

Embedded Inversion in L2 Varieties

Colloquial Indian English is not a L1 dialect but a L2 system acquired after the supposed critical period for syntax (Johnson & Newport, 1989; Long, 2003; Meisel, 2009; Newport & Supalla, 1990), as discussed earlier. The assumption of such a critical period has informed various hypotheses over the years that claim either partial or complete impairment of knowledge of UG following this sensitive window (Clahsen & Muysken, 1989; Hawkins & Casillas, 2008; Hawkins & Chan, 1997; Kanno, 1996; Meisel, 2009; Schachter, 1989; Tsimpli & Dimitrakopoulou, 2007). Second language grammars that purportedly violate universal constraints have been called rogue (Thomas, 1991), illicit (Hamilton, 1998), or wild (Goodluck, 1991). If inversion were truly restricted to embedded clauses in L2 English in India, then it would be a rogue grammar in this sense.

However, as noted previously, I adopt the alternative hypothesis, which is that L2s are natural languages in the same sense as L1s (e.g., Adjemian, 1976; Eckman, 1981; Gass, 2013; Schwartz & Sprouse, 1996; Sharwood Smith, 1988; White, 2003). This is perhaps most succinctly expressed by the interlanguage structural conformity hypothesis (Eckman, Moravcsik, & Wirth, 1989, p. 195):

(22) All universals that are true for primary languages are also true for interlanguages.

Embedded inversion provides a testing ground for this hypothesis, yet surprisingly there appears to have been little systematic investigation of this topic in SLA, despite numerous expressions of theoretical interest, anecdotal reports, and minor experimental reports (e.g., Bley-Vroman, 1997; Borer, 1996; Escutia, 2002; Finegan, 1999; Johnston, 1985). A recent L2 study takes a more systematic look at this. Pozzan and Quirk (2013) investigated inversion in main and embedded clauses by Chinese and Spanish learners of English. Two elicited-production tasks were administered to 80 participants (32 L1 Chinese, 32 L1 Spanish, and 16 L1 English controls). If L1 transfer were involved, one may expect

greater rates of inversion in the case of L1 Spanish (which has T-to-C movement in both main and embedded clauses) than in the case of L1 Chinese (which has no T-to-C movement). The results relevant to the current study are as follows. First, for each L1 group, there was more *wh*-inversion in matrix clauses than in embedded clauses (L1 Spanish, matrix: 85%; L1 Spanish, embedded: 29%; L1 Chinese, matrix: 93%; L1 Chinese, embedded: 29%). This was also true for inversion in *yes*/*no* questions (L1 Spanish, matrix: 97%; L1 Spanish, embedded: 2%; L1 Chinese, matrix: 100%; L1 Chinese, embedded: 2%). Second, the low rates of inversion in embedded *yes*/*no* questions may be plausibly linked to the overt complementizer *if*, which appears to be obligatory for most learners in this study (as in Standard English) and which blocks movement from T to C. Thus, L2 learners appear to respect the same universal principles of phrase structure as L1 dialect speakers. Third, there appears to be no transfer effect (although further studies with lower-level learners are required to confirm this). This study remains in need of replication, but the results are likely to be generalizable and echo the findings from L1 dialect research.

The question of whether *wh*-inversion is restricted to embedded clauses in Indian English is, of course, an empirical one. Although this distribution may seem unlikely in the context of the preceding review, it remains the standard analysis and is in need of systematic evaluation. The implications for L2 research are fundamental: Either colloquial Indian English is an exceptional language variety whose violation of universal principles challenges the notion that L2s are of the same ilk as L1s or, alternatively, it conforms to crosslinguistic constraints and supports the idea that L1s and L2s are equally natural expressions of the human language faculty.

METHOD

The ICE-India corpus was compiled by Shastri (Kolhapur) and Leitner (Berlin); the untagged corpus was released in 2002 and contains 1,000,000 words of spoken and written Indian English from the 1990s. The search tool was a program written in Perl (ActivePerl Version 5.10.0.1004-MSI).[12] Following the original intentions of the ICE project, the participants had all lived their lives in the country where the corpus was collected and had been educated through the medium of English. Although this renders the corpus unrepresentative of the vast cline of English usage in India and excludes consideration of millions of speakers with marginal control of the language, it nevertheless provides a representative snapshot of what Agnihotri (1999) refers to as "fluent users of English in India" (p. 183), who might be considered the arbiters of change if Indian English is a "norm-developing" (B. B. Kachru, 1986/1990, Chapter 5) variety of the language.

The spoken files contain approximately 600,000 words, and the written files 400,000. Due to the attested prevalence of the relevant patterns in speech rather than in writing (Balasubramanian, 2009; Hilbert, 2011; Mesthrie & Bhatt, 2008), searches were restricted to the spoken corpus, which is composed of 300 text files, some of which contain more than 30 pages of single-spaced text.[13] These text files are sorted into four main categories as follows: S1A (100 texts, private dialogues: direct conversations and telephone calls); S1B (80 texts, public dialogues: class lessons, broadcast discussions, broadcast interviews, parliamentary debates, legal cross-examinations, and business transactions); S2A (70 texts, unscripted monologues: spontaneous commentaries, unscripted speeches, demonstrations, and legal presentations); and S2B (50 texts, scripted monologues: broadcast news, broadcast talks, and nonbroadcast talks).

There was a need to further restrict the search, as the spoken corpus contained vast numbers of *wh*-words and auxiliaries, which required meticulous manual filtering; this filtering involved the examination of each utterance in the context of its surrounding discourse. To restrict the results to a number conducive to such analysis, the purpose-designed Perl program targeted particular combinations of *wh*-elements (*what* / *who* / *why* / *when* / *how*); auxiliaries or modals, henceforth AUXs (*am* / *is* / *are* / *was* / *were* / *does* / *do* / *did* / *can* / *could* / *will* / *would* / *have* / *has* / *had*); subject pronouns (*he* / *she* / *they*); and interrogative complementizers (*if* / *whether*). An initial search for these elements in isolation resulted in more than 30,000 hits, but searches for specific combinations resulted in manageable subcorpora. The selection of such elements in a search of this type will always be somewhat controversial, but I assume that the patterns that hold for each of these five *wh*-elements will also hold for others such as *where*, *which*, *whose*, and *whom*, just as the behavior of the 15 selected AUX targets, I assume, can be generalized to elements such as *may*, *might*, *shall*, and *should.* Note that patterns of inversion remain the same (a) across contexts of *wh*-extraction from complements and adjuncts and (b) across types of modals. Third-person pronouns were chosen due to their predominance in indirect speech, as the first and second person are often changed to third person when speech is reported (Henry, 1995).

Following each search conducted, all results were subject to manual filtering, which was fundamental to the analysis. First, the transcription itself does not suffice to identify even main questions. Question mark notation is unreliable, as can be appreciated by anyone who has worked with transcripts of natural dialogue. Although utterance-initial words are capitalized, this does not suffice to distinguish matrix *wh*, as many questions begin with other words, such as *now*, *so*, or *but*. (Perl is case-sensitive, so both upper- and lowercase versions of *wh*-elements were included in searches.) Second, manual filtering at the sentence level is required to distinguish between matrix and embedded *wh*-questions and to eliminate

false positives such as relative clauses and *wh*-comparatives, as mentioned earlier. Third, examination of the discourse before and after the hit is often necessary to determine whether a phrase is, for example, a *wh*-cleft or a noninverted *wh*-question; thus, all hits were analyzed in context.

RESULTS

Main Search: *Wh*+AUX

The purpose of this general search was to identify all instances of inversion with the selected *wh*-elements (*what* / *who* / *why* / *when* / *how*) followed directly by the selected list of AUXs (*am* / *is* / *are* / *was* / *were* / *does* / *do* / *did* / *can* / *could* / *will* / *would* / *have* / *has* / *had*). The main search enabled manual filtering into main and embedded clauses. Parallel identification of cases of noninversion was not possible, as the noun phrase following the *wh*-element in such cases could begin with almost any quantifier, determiner, adjective, or noun. The automatic search returned 2,381 hits, of which 1,539 were manually selected for analysis, and 842 were eliminated. The examples in (23) and (24) show representative instances of unambiguous *wh*-inversion in main clauses and embedded clauses, respectively. In contrast, the examples in (25) illustrate the main types of eliminated utterances: *wh*-clefts, relative clauses, *wh*-comparatives, copied auxiliaries, and subject *wh*-questions.

(23) a. *What is your opinion, about the cultural entity of, our nation.* <S1A-005#57:1:A>[14]
b. *But they, badly require a text book at home, yeah, how will you do the exercise, how will you study.* <S1A-060#152:1:A>
c. *Okay* ahn*, then why did you offer automobile then?* <S1A-061#106:1:B>

(24) a. *Can you tell me what would it depend on?* <S1B-007#73:1:A>
b. *One of my colleagues in the UN retired, uh and I was asking his wife how does she feels now, when the husband is in the home?* <S1B-035#129:1:B>
c. *I don't know why do they have this inhibitions about themselves.* <S1A-062#183:1:A>

(25) a. *But uh, what has happened this year, is really, a political disaster.* <S1A-009#160:1:A> (*wh*-cleft)
b. *A student who is admitted in MA honours in English, he will be exposed to thermodynamics, . . . will be exposed to computer science.* <S1A-023#113:1:A> (relative clause)
c. *No you see we, in Amritsar, there's, it's an extreme weather, I mean it's very cold there now, and it will be hotter than what is here now.* <S1A-013#51:1:B> (*wh*-comparative)
d. *But, what will your poor husband will say if you're so busy with your, studies and other things.* <S1A-097#190:2:B> (copied auxiliary)
e. *I don't know who is interested and not interested.* <S1A-038#269:1:B> (subject *wh*)

The general results for the main search are given in Table 1. In reporting the results, proportions are given in both raw numbers and percentages. Whether inversion or noninversion may be said to be characteristic of main or embedded clauses is indicated not only by descriptive statistics but by binomial tests showing whether the larger of the two proportions is significantly higher than the 50% expected by chance. Effect sizes were also calculated for each percentage's departure from chance according to Cohen's effect size index *g*, in which small = .05, medium = .15, and large = .25 (Cohen, 1988). Cohen's *g* (as compared to Cohen's *d* or Hedges's *g*) was appropriate for the purposes of this analysis because proportions were compared to random chance (for discussion, see Cohen, 1992, p. 99).

Table 1 shows that *wh*-inversion was attested in main clauses at a rate of 1,439/1,539 (93.5%, $p < .001$, $g = 0.435$) and in embedded clauses at a rate of only 100/1,539 (6.5%). This pattern was also found for each individual *wh*-element, as shown in Table 2. Thus it is transparent from the main search that *wh*-inversion is unequivocally a characteristic of matrix clauses.

Subsidiary Search 1: *Wh*+AUX+*Pronoun* / *Wh*+*Pronoun*+AUX

This subsidiary search did not involve a strict subset of the data in the main search; rather, it was an independent, ancillary search. As the main body of data only exemplified patterns of inversion, another, smaller search was conducted specifically to compare rates of inversion and noninversion (in which the *wh*-phrase was followed by some nominal constituent) in matrix and embedded domains. Subsidiary Search 1 involved the three *wh*-words with the greatest number of initial hits (*what*, *how*, and *who*),[15] the same list of AUXs, and three third-person subject pronouns (*he*, *she*, and *they*), as discussed earlier. In this way, rates of inversion and noninversion could be directly compared. In the first part of the search, targeting the pattern of inversion *wh*+AUX+*pronoun*, there were 68 hits, examples of which are given in (26), including the only example of this type in an embedded clause (27).

Table 1. Main search: *Wh*+AUX

Combination	Total	Matrix	Embedded	Eliminated
Wh+AUX	2,381	1,439 (93.5%, $p < .001$, $g = 0.435$)	100 (6.5%)	842

Table 2. Main search: *Wh*+AUX, breakdown by *wh*-type

Combination	Total	Matrix	Embedded	Eliminated
What+AUX	1,192	878 (91.7%, $p < .001$, $g = 0.417$)	79 (8.3%)	235
How+AUX	364	344 (97.2%, $p < .001$, $g = 0.472$)	10 (2.8%)	10
Who+AUX	661	69 (95.8%, $p < .001$, $g = 0.458$)	3 (4.2%)	589
When+AUX	67	66 (100.0%, $p < .001$, $g = 0.500$)	0 (0.0%)	1
Why+AUX	97	82 (91.1%, $p < .001$, $g = 0.411$)	8 (8.9%)	7

(26) a. *How do they want to prevent any damage to the, mosque.* <S1B-038#25:1:B>
b. *What does he realises ultimately.* <S1A-006#60:1:B>
c. *And what does he say, that, he only has, fleeting glimpses that is, vanishing views.* <S1B-011#102:1:A>

(27) *Do you remember that what did she answered, what is the essence.* <S1A-059#39:1:A>

The results for *wh*+AUX+*pronoun* are given in Table 3. This pattern was attested in main clauses at a rate of 67/68 (98.5%, $p < .001$, $g = 0.485$) and in embedded clauses at a rate of only 1/68 (1.5%). The single exception is considered to be no more than noise in the data: It may well have involved a restart following the first complementizer. This distribution was also found for each individual *wh*-element, as shown in Table 4 (although *who* produced only four exemplars). Thus, as expected following the main search, *wh*-inversion was again shown to be robustly associated with main clauses.

In the second part of this subsidiary search, targeting the pattern of noninversion *wh*+*pronoun*+AUX, there were 71 hits, examples of which are given below, for main clauses, shown in (28), and for embedded clauses, shown in (29).

(28) a. *What they will give in breakfast in your college.* <S1A-051#83:1:A>
b. *How they are connected with the rest, of the universe?* <S2B-024#4:1:A>

Table 3. Subsidiary Search 1: *Wh*+AUX+*pronouns*

Combination	Total	Matrix	Embedded
Wh+AUX+*prn*	68	67 (98.5%, $p < .001$, $g = 0.485$)	1 (1.5%)

Note. Prn = pronoun.

Table 4. Subsidiary Search 1: *Wh*+AUX+*pronouns*, breakdown by *wh*-type

Combination	Total	Matrix	Embedded
What+AUX+*prn*	46	45 (97.8%, $p < .001$, $g = 0.478$)	1 (2.2%)
How+AUX+*prn*	18	18 (100.0%, $p < .001$, $g = 0.500$)	0 (0.0%)
Who+AUX+*prn*	4	4 (100.0%, $p < .125$, $g = 0.500$)	0 (0.0%)

Note. Prn = pronoun.

(29) a. *I don't know how they will do it in the next few days.* <S1B-038#58:1:C>
b. *And nobody knows why they are there, and what they are doing.* <S1B-046#130:1:B>

In this case, the search showed that noninversion was not correlated significantly with either main or embedded clauses, as shown in Table 5. The breakdown into *wh*-types can be seen in Table 6. Noninversion with *what* shows a tendency to appear in embedded clauses, which reaches significance. In contrast, noninversion with *how* shows a tendency to appear in main clauses, though this does not reach significance. *Who* produced only one exemplar. Although the appearance of noninversion in main clauses is of a proportion large enough to warrant further study, this pattern is, of course, attested in L1 dialects such as African-American English and has no bearing on the primary question in play: whether *wh*-inversion is restricted to embedded clauses.

In summary, this search revealed an asymmetry in patterns of inversion and noninversion. Within the limited parameters of the search, *wh*-inversion was shown to be very much characteristic of main clauses, with negligible occurrence in embedded clauses. However, noninversion was found in both main and embedded clauses.

Subsidiary Search 2: *If*/*Whether*+AUX

As mentioned earlier, *whether* and *if* are both considered to be interrogative complementizers for current purposes. Although *whether* is often analyzed as a type of *wh*-element, it is in complementary distribution

Table 5. Subsidiary Search 1: *Wh*+*pronouns*+AUX

Combination	Total	Matrix	Embedded
Wh+*prn*+AUX	71	29 (40.8%)	42 (59.2%, $p = .154$, $g = 0.092$)

Note. Prn = pronoun.

Table 6. Subsidiary Search 1: *Wh+pronouns*+AUX, breakdown by *wh*-type

Combination	Total	Matrix	Embedded
What+prn+AUX	49	16 (32.7%)	33 (67.3%, p = .021, g = 0.173)
How+prn+AUX	21	13 (61.9%, p = .383, g = 0.119)	8 (38.1%)
Who+prn+AUX	1	0 (0.0%)	1 (100.0%, p = 1.000, g = 0.5)

Note. Prn = pronoun.

with the complementizer *if* in all environments involving the embedding of finite clauses (see Note 9): Both elements absolutely prohibit inversion in all L1 dialects. As all varieties of English have interrogative complementizers, it cannot be the case that there exists a grammar that makes inversion obligatory in all embedded clauses. The prediction that follows for L2 English in India is that although inversion may optionally occur following movement of *wh*-elements to Spec, CP, and although it may be obligatory following complementizer deletion in *yes*/*no* questions, it should be impossible following an overt interrogative complementizer. The purpose of Subsidiary Search 2 was to identify all patterns of inversion and noninversion with *if*/*whether* with the same list of AUXs and the same list of pronouns as used in Subsidiary Search 1. All clauses were considered to be embedded, whether they followed the main clause or whether they preceded it as a result of a fronting operation (true main clause operations, such as tag questions, never operate in *if*/*whether* clauses). There were 60 hits in total for *if* and 41 for *whether* within the constraints of the search. Only three instances of inversion were found. Examples (30) and (31) illustrate inversion and noninversion, respectively. The main clause supporting the *whether* clause in (30b) appears earlier in the discourse: "You should ensure . . ." <S2B-029#29:1:A>.

(30) a. *At this age if is he going to have any surgery.* <S1A-037#82:1:A>[16]
b. *Or if you take the examples of domestic electrical appliances, whether do they carry, I.S.I. mark, given by, bureau of Indian standards in the country.* <S2B-029#30:1:A>

(31) a. *If they can be called as Indo-Europeans, then what is the, need for the use of word Indo, uh uh Indo-Aryans?* <S1B-005#158:1:A>
b. *Now he wanted to know, uh the forty-seven children studying in Kannada, uh studying Kannada in English medium school whether they are outsiders or, uh insiders.* <ICE-IND:S1B-080#147:1:B>

The general results for *if/whether*+AUX+*pronoun* and *if/whether*+*pronoun*+AUX are given in Table 7. Noninversion is robustly attested, for *if* at a rate of 59/60 (98.3%, p < .001, g = 0.483) and for *whether* at a rate

Table 7. Subsidiary Search 2: *If/whether*+AUX+*pronouns* and *if/whether+pronouns*+AUX

Combination	Total	AUX+*prn*	*prn*+AUX
If	60	1 (1.7%)	59 (98.3%, $p < .001$, $g = 0.483$)
Whether	41	2 (4.9%)	39 (95.1%, $p < .001$, $g = 0.451$)

Note. Prn = pronoun.

of 39/41 (95.1%, $p < .001$, $g = 0.451$). Although this corpus study does not encompass data from L1 varieties, it is to be expected that occasional instances of inversion may surface at similar rates in monolingual speakers and are presumably not reflective of underlying grammatical competence. This subsidiary search confirms that, in the environment of overt interrogative complementizers, inversion is ruled out, such that speakers of L2 English in India conform to the same constraints observed by monolingual speakers of L1 dialects, reflecting unconscious knowledge of universal principles of grammar.

DISCUSSION

The results clearly demonstrate that the traditional account of *wh*-inversion in Indian English grammar cannot be maintained, despite recent attempts to formalize this in terms of syntactic constraint ranking. It is not the case that inversion is restricted to embedded clauses and ruled out in main clauses. Rather, both inverted and noninverted forms occur in matrix interrogatives, such that neither can be said to be obligatory; in this sense, inversion may have a similar range of forms to those found in African-American English (L. J. Green, 2002, pp. 85–86). In embedded clauses, far from being compulsory, *wh*-inversion was shown to occur in only a very small proportion of utterances, such that it can unequivocally be described as primarily a main clause phenomenon. The vast majority of these cases involved question verbs or interrogative or negative resolutive verbs, although there were some cases of affirmative resolutive verbs, similar to the Belfast English examples in (12). Moreover, inversion in embedded clauses was limited to syntactic structures in which the C node was able to serve as a landing site: Overt expression of *if*/*whether* clearly blocks movement. These results match the initial reports on embedded inversion in the corpus linguistics literature and add inferential statistical support for the distribution.

Given the clarity of the results, a natural question that arises is why the conventional account of interrogatives in Indian English has been so widely accepted for so many years. Again, an acquisitionist perspective might be

useful here. If one asks parents if their children often overgeneralize *-ed* to irregular past tense forms (e.g., *I goed to the park; She drawed a picture; The bird flied away*), the answer is likely to be, "Yes, all the time!" However, when Marcus et al. (1992) examined transcriptions of spontaneous speech from 83 children aged 1 year and 3 months to 6 years and 6 months and extracted 11,500 sentences with irregular past tense forms, they found that such forms represented a small minority for each child at all stages of development, the average error rate being only 4%. The common impression of their ubiquity stems from increased noticing of nontargetlike forms. It is possible that the proliferation of anecdotal reports that consolidated the accepted account of embedded inversion in Indian English arose in a similar way—through increased noticing by highly educated observers of forms that do not conform to prescriptive standards.

The phenomenon of copied auxiliaries and modals merits further comment. As noted earlier, Hilbert (2011) argues that in embedded *wh*-questions, inversion is restricted to the verb *be* and essentially involves chunking of the *wh*-element with the third-person singular *is*. She suggests that the co-occurrence of *wh+is* with a copy (as in the Singapore English example *Don't understand how's it is coming up*) is a "fundamental aspect of support for the 'fixed chunk' analysis of what otherwise looks like inversion" (p. 132). However, none of the examples given in (24) involve the verb *be*, so they cannot possibly be explained by the chunking analysis. Moreover, examples such as (25d) without *be* but with copied auxiliaries also resist a chunking analysis, although they follow straightforwardly from a standard generative account in which an element moves to a higher position in the syntax, leaving a copy in the base position. In the acquisition literature, copied auxiliaries are linked to other phenomena also present in the current dataset, such as double agreement marking (as in [24b] and [26b]) and double tense marking (as in [27]). These cases involve verbs other than *be* and include *do*-insertion; they imply not an analysis based on imitation but rather universal patterns of transformational syntax. It should also be noted that, despite the "noticing effect," such cases are very rare in the data, as may be expected: Stromswold (1990) reported a frequency of doubling errors of only 0.4 % in question structures involving inversion in L1 acquisition, which suggests that this may be due to processing difficulty in real-time production rather than a discrete stage of acquisition.

One possible objection to the generalization of these results is that the ICE-India corpus is representative not of colloquial Indian English but rather of an educated variety syntactically closer to international standards. However, as noted earlier, if Indian English is a norm-developing variety of the language, then the arbiters of change and the setters of new standards are precisely people communicating in these contexts in which English is used for intranational purposes:

in the realms of public education, business transactions, courtrooms, political debates, and broadcast media. That this nationally accepted variety is nevertheless a L2 for most speakers is revealed in the corpus not only by forms that have stabilized but by other forms of the type discussed previously, which are not considered part of the emerging standard grammar of Indian English.

The fact remains that there is a cline of bilingualism in India. Verma (1980) describes this cline as "extending from non-educated varieties of English at one end (which are not all intelligible) to an internationally accepted standard form at the other" and notes that "in between these two ends we have a great range of variation" (p. 83). On this broadly accepted view, standardizing Indian English is situated "somewhere at the central point of the cline" (Gupta, 2001, p. 156). What is important to understand is that one cannot locate a uniform variety termed "colloquial Indian English" at some putative point at the lower end of the cline. Speakers who have more marginal control of the language come from considerably diverse linguistic backgrounds, which feature hundreds of L1s, including Indo-Aryan languages such as Hindi/Urdu, Bengali, Punjabi, and Marathi; Dravidian languages such as Kannada, Malayalam, Tamil, and Telugu; and other languages from the Austro-Asiatic and Tibeto-Burman families. This linguistic diversity must necessarily be reflected in diverse interlanguages. The lack of information about L1s is a major limitation in most World Englishes corpora; thus, studies based on such corpora must be supplemented with the kind of experimental work that allows for greater individual analysis of the linguistic code repertoires of participants.

From a generative L2 perspective, the investigation of proposed language universals can be extended across all combinations of L1s and L2s. Yet interest in commonalities across varieties in the World Englishes literature has generally been narrowly confined to the documentation of "Angloversals" (Mair, 2003) with reference to other postcolonial varieties (for descriptive lists of proposed Angloversals, see Mesthrie & Bhatt, 2008, and Schneider, 2007; for justification of the term itself, see Szmrecsanyi & Kortmann, 2009). This English Studies perspective could be significantly informed from the vantage point of L2 acquisition in general. For example, it has been known at least since Platt, Weber, and Ho (1984) that fluctuation between definite and specific interpretations of the definite article has characterized several World English varieties, yet this remains to be investigated from a theoretical L2 perspective such as that found in Ionin, Ko, and Wexler (2004). Similarly, variable omission of third-person agreement and past tense *-ed* is much remarked on, but insights may be gained if investigations were informed by L2 studies such as Prévost and White (2000). As for *wh*-inversion as a main clause phenomenon, the results of this study show that L2 English in India respects universal constraints, despite widely

accepted accounts to the contrary. This should be expected if L2 grammars are just as systematic and logical as L1 grammars. These findings support the notion that L2 systems functioning as part of larger code repertoires may be investigated with the same theoretical tools as used for monolingual native speaker varieties.

As a final observation, it should be noted that the population of relevance to the current study is woefully underrepresented in mainstream SLA. Most research on L2 English has focused on classroom learners, either in study-abroad situations in the United States or the United Kingdom or in foreign language classes with an emphasis on international communication, whereas the potential of many other significant populations of L2 users has gone largely untapped. There are approximately the same number of L2 users of English in postcolonial societies in which English is used for intranational communication as there are L1 users of English in the world. Indeed, these particular populations of regular L2 users may soon outnumber L1 users (Crystal, 2006). Yet they are essentially absent from the L2 literature. If this is to be redressed, then expanding L2 research in this direction involves a continuing shift in conceptualization of the L2 acquisition process (in line with earlier work by Cook, 1991, 2002, 2003, and others). To reiterate the observations made at the outset of this article, it is fruitless in such contexts to invoke the notion of a target language spoken by idealized native speakers. In influential work by Abrahamsson and Hyltenstam (2009), the idea of a native speaker is implicitly equated with the idea of having the same competence as a monolingual. However, if this interpretation were valid, we would be forced to reach the absurd conclusion that there are almost no native speakers of any language in the major urban centers of the world such as Accra, Delhi, Nairobi, Kuala Lumpur, Lagos, Lahore, Manila, or Cape Town, where multilingualism is a linguistic reality. Such contexts provide a superabundance of languages in contact with L2 English and constitute an extremely rich environment in which to explore current issues of L2 acquisition and knowledge.

CONCLUSION

For many years, the conventional account of *wh*-inversion in L2 English in India has been that it is the mirror opposite of standard varieties in the United Kingdom and United States, with no inversion in main clauses and obligatory inversion in embedded clauses. This claim was given formal expression in the OT analysis proposed by Bhatt (2000) and Mesthrie and Bhatt (2008), which restricted inversion to embedded clauses by means of syntactic constraint ranking. The current study challenged this account on both theoretical and empirical grounds. In line with recent observations in corpus linguistic research, and in

accordance with the interlanguage structural conformity hypothesis (Eckman et al., 1989), a search of the ICE-India corpus revealed that Indian English conforms to the same principles of phrase structure and the same universal constraints as found in other L1 and L2 varieties of English. *Wh*-inversion is primarily a characteristic of main clauses and is possible in embedded clauses only if the appropriate syntactic and pragmatic conditions obtain. More broadly, this supports a view of multilingual systems as internally coherent and subject to the constraints of UG. It is to be hoped that the current study might serve to stimulate an expansion of the horizons of formal L2 research to include a largely neglected yet extremely rich linguistic environment: that of L2 varieties of English in postcolonial societies.

Received 26 February 2013
Accepted 9 December 2013
Final Version Received 19 February 2014

NOTES

1. Following convention, the abbreviation *SLA* is used to refer the field, whereas L2 is preferred for other uses—for example, L2 acquisition, L2 knowledge, and L2 research.

2. The observation also applies to *yes*/*no* questions, but the focus of this study is specifically *wh*-questions. The term *auxiliary inversion* (or simply *inversion*) is used throughout to facilitate comprehension across syntactic frameworks and across the disciplines of SLA and World Englishes. The analysis adopted here is not one of inversion of elements but rather the standard analysis in mainstream generative syntax, in which the auxiliary or modal moves from T (tense) to C (complementizer) position (i.e., T-to-C movement).

3. By way of comparison, the number of students enrolled in intensive English programs in the United States in the academic year 2011–2012 was 38,887 (Institute of International Education, 2012).

4. These census data were drawn from a synthesis of the official data made by Nagle (2010), which was based on various online and print publications of the Office of the Registrar General and Census Commissioner, Government of India.

5. It is the colloquial variety that I am here equating with L2 English in India, as the standard is identical in the relevant respects to other Standard Englishes around the world. It must be noted that, although colloquial Indian English is clearly a L2 according to definitions in the field of SLA as discussed earlier, it is considered impolitic and sometimes even unethical in the World Englishes literature to make a distinction between L1 and L2 in multilingual societies of this type. Many who have become L2 dominant would claim English as a L1, even though it was learned in late childhood or adulthood and even though the phonological, syntactic, and lexical influence from L1s in colloquial Indian English is uncontroversial.

6. See Bhatt (2000), Appendix, tokens 3, 7, 10, 11, 19, 22.

7. It should be noted that Joan Bybee (Bybee, 2002) and Joan Hooper (Hooper & Thompson, 1973) are the same person.

8. For recent developments in the cartographic approach to clausal projections, along the lines of Rizzi (1997), see Aelbrecht et al. (2012) and Haegeman (2012).

9. For the purposes of this article, *if* and *whether* are both described as interrogative complementizers, as they have exactly the same distribution in the relevant syntactic environments. However, there is some evidence that *whether* is hosted not in the head but in the Spec of the CP. Haegeman and Gueron (1999, pp. 175–176) observe that heads but

not specifiers place selectional restrictions on complements. *If*, like *that*, selects a finite complement, but *whether*, like other *wh*-words, does not select for finiteness:

(i) *I wonder if I should go.* / **I wonder if to go.*
(ii) *I expected that I would go.* / **I expected that to go.*
(iii) *I can't decide whether I should go.* / *I can't decide whether to go.*
(iv) *I wonder where I should go.* / *I wonder where to go.*

Nevertheless, a simple *whether*-in-Spec, CP analysis cannot explain why the head C is blocked as a landing site only in the case of *whether* and not other *wh*-elements:

(v) *I asked* {*where* / *when* / *why* / **whether* / **if*} *should I go.*

10. For clarity, it should be noted that McCloskey (2006) does not adopt a simple embedded clause analysis but rather uses arguments from adjunction phenomena to posit a double-headed CP structure, discussion of which remains outside the scope of this article.

11. *Gan* = go, *me* = my.

12. The particular version was ActivePerl (Version 5.10.0.1004–MSI); for this task, Perl is consistent across installations. In February 2013, tagged ICE corpora became officially available for the first time, tagged with CLAWS7 (Garside, 1987) and the USAS semantic tagger (Wilson & Rayson, 1993), which will be useful for future investigations. Purpose-designed search programs using Perl have the additional advantage of applicability to both tagged corpora (to query them) and nontagged corpora (including any online transcripts in the public domain).

13. Readers familiar with the huge English language corpora available in the United States or the United Kingdom might find the 1-million-word ICE corpora relatively small. By comparison, the Corpus of Contemporary American English (COCA; Davies, 2008) now has 450 million words, and the British National Corpus (2007) has 100 million words. However, the quantity of data is more than sufficient for statistical purposes, and ICE represents a major breakthrough in making World Englishes more accessible for linguistic research.

14. All corpus examples have a simplified markup for present purposes. Pauses are represented with simple commas. Indigenous expressions have no markup but are, instead, presented in italics and remain unglossed, as they are irrelevant for the analysis. Certain strings that make no sense in isolated examples, such a back-channeling in mid-sentence during overlapping speech, have been replaced by [. . .], as have extended parts of utterances that have no relevance to the analysis. Question marks are only used when they were part of the original markup.

15. Ultimately, there were more examples of inversion with *why* than with *who* following manual filtering, as 589/661 examples with *who* were eliminated, as compared to 7/97 with *why*. Eliminated *who* utterances involved many relative clauses as well as subject *wh*-questions and *wh*-clefts (see examples [25a], [25b], and [25e]).

16. An anonymous reviewer correctly points out that these examples involve "conditional *if*," which might be distinct from *if* as an interrogative complementizer. In the current analysis, *if* and *whether* are given uniform treatment in such cases, which are assumed to involve the fronting of an embedded clause and are considered examples of "devices which determine nonveridical contexts" (McCloskey, 2006, p. 114), as discussed previously. The truth value of the clause is questioned by the complementizer in such cases, which seems relevant in considering the source of what I assume to be performance errors, as in (30).

REFERENCES

Abrahamsson, N., & Hyltenstam, K. (2009) Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*, 249–306.
Adjemian, C. (1976). On the nature of interlanguage systems. *Language Learning*, *26*, 297–320.
Aelbrecht, L., Haegeman, L., & Nye, R. (2012). *Main clause phenomena: New horizons*. Amsterdam, the Netherlands: Benjamins.

Agnihotri, R. K. (1999). Regional report on India: English in India. In R. Singh (Ed.), *The yearbook of South Asian languages and linguistics* (pp. 161–183). London, UK: Sage.
Balasubramanian, C. (2009). *Register variation in Indian English*. Amsterdam, the Netherlands: Benjamins.
Bhatt, R. M. (2000). Optimal expressions in Indian English. *English Language and Linguistics*, *4*, 69–95.
Bhatt, R. M. (2008). Indian English: Syntax. In R. Mesthrie (Ed.), *Varieties of English: Vol. 4. Africa, South and South-East Asia* (pp. 546–562). Berlin, Germany: Mouton de Gruyter. (Reprinted from *A handbook of varieties of English: Vol. 2. Morphology and syntax*, pp. 1016–1030, by B. Kortmann, E. W. Schneider, K. Burridge, R. Mesthrie, & C. Upton, Eds., 2004, Berlin, Germany: Mouton de Gruyter).
Bley-Vroman, R. (1997, October). *Features and patterns in foreign language learning*. Plenary presented at the Second Language Research Forum, Michigan State University, East Lansing.
The British National Corpus (Version 3; BNC XML Edition) [Electronic corpus]. (2007). Retrieved from http://www.natcorp.ox.ac.uk/
Borer, H. (1996). Access to Universal Grammar: The real issues. Commentary on Epstein, Flynn & Martohardjono. *Brain and Behavioral Sciences*, *19*, 718–720.
Bybee, J. L. (2002). Main clauses are innovative, subordinate clauses are conservative: Consequences for the nature of constructions. In J. L. Bybee & M. Noonan (Eds.), *Complex sentences in grammar and discourse: Essays in honor of Sandra A. Thompson* (pp. 1–17). Philadelphia, PA: Benjamins.
Clahsen, H., & Muysken, P. (1989). The UG paradox in L2 acquisition. *Second Language Research*, *5*, 1–29.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.
Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*, 98–101.
Cook, V. J. (1991). The poverty-of-the-stimulus argument and multi-competence. *Second Language Research*, *7*, 103–117.
Cook, V. J. (Ed.). (2002). *Portraits of the L2 user*. Clevedon, UK: Multilingual Matters.
Cook, V. J. (Ed.). (2003). *Effects of the second language on the first*. Clevedon, UK: Multilingual Matters.
Crystal, D. (2006). English worldwide. In R. Hogg & D. Denison (Eds.), *A history of the English language* (pp. 420–439). Cambridge, UK: Cambridge University Press.
Davies, M. (2008). The Corpus of Contemporary American English: 450 million words, 1990–present [Electronic corpus]. Retrieved from http://corpus.byu.edu/coca/
Dekydtspotter, L. (2009). Second language epistemology, take two. *Studies in Second Language Acquisition*, *31*, 291–321.
Eckman, F. (1981). On the naturalness of interlanguage rules. *Language Learning*, *31*, 195–216.
Eckman, F., Moravcsik, E., & Wirth, J. (1989). Implicational universals and interrogative structures in the interlanguage of ESL learners. *Language Learning*, *39*, 173–205.
Emonds, J. E. (1970). *Root and structure-preserving transformations* (Unpublished doctoral dissertation). MIT, Cambridge, MA.
Emonds, J. E. (1976). *A transformational approach to English syntax: Root, structure-preserving and local transformations*. New York, NY: Academic Press.
Emonds, J. E. (2004). Unspecified categories as the key to root constructions. In D. Adger, C. De Cat, & G. Tsoulas (Eds.), *Peripheries: Syntactic edges and their effects* (pp. 75–120). Dordrecht, the Netherlands: Kluwer.
Emonds, J. E. (2012). Augmented structure preservation and the Tensed S Constraint. In L. Haegeman, R. Nye, & L. Aelbrecht (Eds.), *Main clause phenomena: New horizons* (pp. 21–46). Amsterdam, the Netherlands: Benjamins.
Escutia, M. (2002, May). *Spanish learners' production of English embedded interrogative clauses*. Paper presented at Circulo de Linguistica Aplicada a la Comunicación (CLAC), Universidad Complutense de Madrid, Spain.
Finegan, E. (1999). *Language: Its structure and use*. Fort Worth, TX: Harcourt.
Garside, R. (1987). The CLAWS word-tagging system. In R. Garside, G. Leech, & G. Sampson (Eds.), *The Computational Analysis of English: A Corpus-Based Approach* (pp. 30–41). London, UK: Longman.

Gass, S. M. (with Behney, J., & Plonsky, L.) (2013). *Second language acquisition: An introductory course* (4th ed.). New York, NY: Routledge.
Goodluck, H. (1991). *Language acquisition: A linguistic introduction*. Oxford, UK: Blackwell.
Green, G. M. (1996). *Pragmatics and natural language understanding*. Mahwah, NJ: Erlbaum.
Green, L. J. (2002). *African American English: A linguistic introduction*. New York, NY: Cambridge University Press.
Greenbaum, S. (Ed.). (1996). *Comparing English worldwide: The International Corpus of English*. Oxford, UK: Clarendon Press.
Grimshaw, J. (1997). Projections, heads, and optimality. *Linguistic Inquiry*, *28*, 373–422.
Gupta, R. S. (2001). English in post-colonial India: An appraisal. In B. Moore (Ed.), *Who's centric now? The present state of post-colonial Englishes* (pp. 148–164). Oxford, UK: Oxford University Press.
Haegeman, L. (2012). *Adverbial clauses, main clause phenomena, and composition of the left periphery: The cartography of syntactic structures*. Oxford, UK: Oxford University Press.
Haegeman, L., & Gueron, J. (1999). *English grammar: A generative perspective*. Oxford, UK: Blackwell.
Hamilton, R. (1998). Underdetermined binding of reflexives by adult Japanese-speaking learners of English. *Second Language Research*, *14*, 292–320.
Hawkins, R., & Casillas, G. (2008). Explaining frequency of verb morphology in early L2 speech. *Lingua*, *118*, 595–612.
Hawkins, R., & Chan, C. Y. (1997). The partial availability of Universal Grammar in second language acquisition: The "failed functional features hypothesis." *Second Language Research*, *13*, 187–226.
Henry, A. (1995). *Belfast English and standard English: Dialect variation and parameter setting*. Oxford, UK: Oxford University Press.
Hilbert, M. (2008). Interrogative inversion in non-standard varieties of English. In P. Siemund & N. Kitana (Eds.), *Language contact and contact languages*. Amsterdam, the Netherlands: Benjamins.
Hilbert, M. (2011). Interrogative inversion as a learner phenomenon in English contact varieties: A case of Angloversals? In J. Mukherjee & M. Hundt (Eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap* (pp. 125–144). Amsterdam, the Netherlands: Benjamins.
Hooper, J. B., & Thompson, S. A. (1973). On the applicability of root transformations. *Linguistic Inquiry*, *4*, 465–497.
Institute of International Education. (2012). International students by field of study, 2010–12. *Open doors report on international educational exchange*. Retrieved from Institute of International Education website: http://www.iie.org/Research-and-Publications/Open-Doors/Data/International-Students/Fields-of-Study/2010-12
Ionin, T., Ko, H., & Wexler, K. (2004). Article semantics in L2-acquisition: The role of specificity. *Language Acquisition*, *12*, 3–69.
Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99.
Johnston, M. (1985). *Syntactic and morphological progressions in learner English*. Canberra, Australia: Department of Immigration and Ethnic Affairs.
Kachru, B. B. (1985). The sacred cows of English. *English Today*, *16*, 3–8.
Kachru, B. B. (1990). *The alchemy of English: The spread, functions and models of non-native English*. Urbana: University of Illinois Press. (Reprinted from *The alchemy of English: The spread, functions and models of non-native English*, by B. B. Kachru, 1986, Oxford, UK: Pergammon Press).
Kachru, B. B. (2006). World Englishes and culture wars. In B. B. Kachru, Y. Kachru, & C. L. Nelson (Eds.), *The handbook of World Englishes* (pp. 446–472). Oxford, UK: Blackwell.
Kachru, Y. (1994). Monolingual bias in SLA research. *TESOL Quarterly*, *28*, 795–800.
Kanno, K. (1996). The status of a non-parametrized principle in the L2 initial state. *Language Acquisition*, *5*, 317–332.
Labov, W., Cohen, P., Robbens, C., & Lewis, J. (1968). *A study of the non-standard English of Negro and Puerto-Rican speakers in New York City: Vol. 1. Phonological and grammatical analyses*. New York, NY: Columbia University.

Lightfoot, D. (2012). Explaining matrix/subordinate domain discrepancies. In L. Haegeman, R. Nye, & L. Aelbrecht (Eds.), *Main clause phenomena: New horizons* (pp. 159–176). Amsterdam, the Netherlands: Benjamins.

Long, M. H. (2003). Stabilization and fossilization in interlanguage development. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 487–535). Oxford, UK: Blackwell.

Mair, C. (2003). *Kreolismen und verbales Identitätsmanagement im geschriebenen jamaikanischen Englisch* [Creolisms and verbal identity management in written Jamaican English]. In E. Vogel, A. Napp, & W. Zwischen (Eds.), *Ausgrenzung und Hybridisierung* (pp. 79–96). Würzburg, Germany: Ergon.

Marcus, G., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in language acquisition [Monograph]. *Monographs of the Society for Research in Child Development, 57*(4, Serial No. 228), v–165. Chicago, IL: University of Chicago Press.

McCloskey, J. (2006). Questions and questioning in a local English. In R. Zanuttini, H. Campos, E. Herburger, & P. H. Portner (Eds.), *Crosslinguistic research in syntax and semantics: Negation, tense, and clausal architecture* (pp. 87–126). Washington, DC: Georgetown University Press.

Meisel, J. M. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft*, *28*, 5–34.

Mesthrie, R., & Bhatt, R. M. (2008). *World Englishes: The study of new linguistic varieties*. Cambridge, UK: Cambridge University Press.

Montrul, S. (2004). *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and in adult L2 acquisition*. Amsterdam, the Netherlands: Benjamins.

Mukherjee, J., & Hundt, M. (Eds.). (2011). *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap*. Amsterdam, the Netherlands: Benjamins.

Nagle, T. (2010). *Language maintenance, language shift, diglossia, bilingualism? The changing roles of English in post-independence India*. Unpublished manuscript, Department of Linguistics, Indiana University, Bloomington.

Newport, E. L., & Supalla, T. (1990). *A critical period effect in the acquisition of a primary language*. Unpublished manuscript, Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY.

Ortega, L. (2005). For what and for whom is our research? The ethical as transformative lens in instructed SLA. *Modern Language Journal*, *89*, 427–443.

Ortega, L. (2013). SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning*, *63*(Suppl. 1), 1–24.

Platt, J., Weber, H., & Ho, M. L. (1984). *The New Englishes*. London, UK: Routledge & Kegan Paul.

Pozzan, L., & Quirk, E. (2013). Second language acquisition of English questions: An elicited production study. *Applied Psycholinguistics*. Advance online publication. doi:10.1017/S0142716412000690

Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, *16*, 103–133.

Prince, A., & Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar*. Unpublished manuscript, Department of Linguistics, Rutgers University, New Brunswick, NJ, and Computer Science Department, University of Colorado at Boulder.

Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar* (pp. 281–337). Dordrecht, the Netherlands: Kluwer.

Rizzi, L. (2006). On the form of chains: Criterial positions and ECP effects. In L. Cheng & N. Corver (Eds.), *Wh-movement: Moving on* (pp. 97–133). Cambridge, MA: MIT Press.

Ross, J. R. (1973). The Penthouse Principle and the order of constituents. In C. W. Corum, T. C. Smith-Stark, & A. Weiser (Eds.), *You take the high node and I'll take the low node* (pp. 397–422). Chicago, IL: Chicago Linguistic Society.

Schachter, J. (1989). Testing a proposed universal. In S. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 73–88). Cambridge, UK: Cambridge University Press.

Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. Cambridge, UK: Cambridge University Press.
Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, *12*, 40–72.
Scott, M. (2004). *WordSmith Tools version 4.0*. Oxford, UK: Oxford University Press.
Sedlatschek, A. (2009). *Contemporary Indian English: Variation and change*. Amsterdam, the Netherlands: Benjamins.
Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, *10*, 209–230.
Sharwood Smith, M. (1988). On the role of linguistic theory in explanations of second language developmental grammars. In S. Flynn & W. O'Neil (Eds.), *Linguistic theory in second language acquisition* (pp. 173–198). Dordrecht, the Netherlands: Kluwer.
Singh, R. (Ed.). (1998). *The native speaker: Multilingual perspectives*. London, UK: Sage.
Slabakova, R. (2008). *Meaning in the second language*. Berlin, Germany: De Gruyter Mouton.
Sridhar, K. K., & Sridhar, S. N. (1986). Bridging the paradigm gap: Second language acquisition theory and indigenized varieties of English. *World Englishes*, *5*, 3–14.
Stromswold, K. (1990). *Learnability and the acquisition of auxiliaries* (Unpublished doctoral dissertation). MIT, Cambridge, MA.
Szmrecsanyi, B., & Kortmann, B. (2009). Vernacular universals and angloversals in a typological perspective. In M. Filppula, J. Klemola, & H. Paulasto (Eds.), *Vernacular universals and language contacts: Evidence from varieties of English and beyond* (pp. 33–53). London, UK: Routledge.
Thomas, M. (1991). Do second language learners have "rogue" grammars of anaphora? In L. Eubank (Ed.), *Point counterpoint: Universal Grammar in the second language* (pp. 375–388). Amsterdam, the Netherlands: Benjamins.
Trudgill, P., & Hannah, J. (2008). *International English: A guide to the varieties of standard English* (5th ed.). Oxford, UK: Oxford University Press.
Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from *wh*-interrogatives in second language acquisition. *Second Language Research*, *23*, 215–242.
Unsworth, S. (2005). *Child L2, adult L2, child L1: Differences and similarities. A study on the acquisition of direct object scrambling in Dutch*. Utrecht, the Netherlands: LOT Dissertation Series.
Verma, S. K. (1980). Swadeshi English: Form and function. *Indian Linguistics*, *41*, 73–84.
White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge, UK: Cambridge University Press.
Whitworth, G. C. (2007). *Indian English: An examination of the errors of the idiom made by Indians in writing English*. In K. Bolton & B. B. Kachru (Eds.), *Asian Englishes* (pp. 1–255). New York, NY: Routledge. (Reprinted from *Indian English: An examination of the errors of the idiom made by Indians in writing English*, by G. C. Whitworth, 1907, Letchworth, UK: Garden City Press).
Wilson, A., & Rayson, P. (1993). Automatic content analysis of spoken discourse. In C. Souter & E. Atwell (Eds.), *Corpus based computational linguistics* (pp. 215–226). Amsterdam, the Netherlands: Rodopi.