

EMBEDDED INVERSION WORLDWIDE¹

Daniela Kolbe and Andrea Sand

Abstract

The inversion of (auxiliary) verb and subject in subordinate interrogative clauses (embedded inversion, or EI) is a feature that occurs in many non-standard varieties of English, especially in varieties that have developed in language contact situations, such as Irish English, East African English or Indian English. Various sources of origin have been proposed in previous research, among them substrate influence or transfer errors of learners in language contact situation. This paper introduces the phenomenon in question and provides an overview of previous research. Drawing on data of the International Corpus of English (ICE), it then presents the results of probabilistic statistic analyses (logistic regression) in order to identify which external and internal factors are strongest in eliciting the inverted word order.

Keywords: *syntax, non-standard syntax, worldwide varieties of English, reported speech, indirect questions, inversion*

1. Introduction

Embedded inversion (EI) is the inversion of (auxiliary) verb and subject in embedded (subordinate) interrogative clauses, sometimes also referred to as indirect questions, as in a).

a) *Jan came to me and asked **would it** be okay if she were to do a recording (unscripted speech, Ireland)*

In Standard English, the inversion of verb and subject is grammatical only in direct questions, as b).

b) *Jane asked me, "Would it be okay if I were to do a recording?"*

¹ This paper is a revised version of a paper presented at the first ISLE Conference at the University of Freiburg (Germany) in October 2008. We would like to thank the discussants of our paper and especially Michaela Hilbert for their valuable input.

The Standard English equivalent of a) would be:

c) *Jane asked me if it would be okay if she was to do a recording.*

with an additional complementizer *if*.

Embedded inversion as displayed in a) occurs in many varieties of English around the world. It is reported in the description of thirteen varieties from four continental regions in Kortmann et al. (2004), e.g., in Irish English (Filppula 2004: 93-95), colloquial American English (Murray and Simon 2004: 224), East African English (Schmied 2004: 936) and Indian English (Bhatt 2004: 1020).

Despite its regional pervasiveness, this syntactic structure seems unusual enough to invite explanations for its existence. These explanations range from speakers' disability to construct correct sentences (e.g., Miller and Weinert 1998: 83) to its existence in Standard English even in literary style under certain circumstances (Denison 1998: 245-246).

The aim of this study is to compare the use of embedded inversion in different varieties of English on the one hand and to test various hypotheses with regard to its origin and distribution. In section 2, we introduce previous research on embedded inversion, discussing both its potential origin and its regional diffusion. Section 3 contains an overview of the data this study draws on and section 4 discusses methodology. Section 5 then presents the results of our analyses which are further discussed in our conclusions in section 6.

2. Previous Research

As mentioned above, embedded inversion is listed among the non-standard syntactic features of many varieties around the world. Many authors have also discussed possible sources for the occurrence of EI. These will be presented below.

2.1. Embedded Inversion in Varieties of English

Of the varieties of English spoken on the British Isles, embedded inversion is included in the treatments of Irish English syntax by Filppula (2004: 93-95) and Harris (1993: 168). It is mentioned in the descriptions of Scottish English by Catford (1957: 110), Sabban (1982: 460-483) and Miller (1993: 126) and Welsh English (Penhallurick 1991: 209-210; Parry 1999: 119; Thomas 1997: 79). Notably, these three varieties of English are all "Celtic Englishes" (cf. Tristram 1997).

However, embedded inversion also occurs in the descriptions of many so-called New Englishes which have developed in former colonies. Platt, Weber and Ho (1984: 127-128) and Simo Bobda (1998: 10) consider it to be a feature of New Englishes in general, while Schmied (2008: 459) reports it as a feature of African Englishes. It is also reported as a feature of Indian English (Sridhar 1992: 144-145, 1996: 10), Jamaican English (Sand 1999: 142-143) and Singapore English (Tay 1982: 63).

Biber et al (1999: 920-921) consider embedded inversion an informal “compromise between direct and indirect speech” that occurs in general colloquial English. It thus appears to be a feature that can be seen as both regional and non-standard.

2.2. Explanations Given for Embedded Inversion

All varieties of English in the descriptions of which embedded inversion is discussed have developed in language contact situations. Hence, the inversion of auxiliary and subject is often claimed to originate in the (or one of the) substrate languages, e.g., as put forth by Filppula (2000) for varieties of English with Celtic substrate, e.g., Irish English, and by Sridhar (1992: 144-145) for Indian English. Sridhar points out that in Dravidian languages, speakers tend not to use reported speech but rather to use a direct quote introduced by a quotative particle which functions similarly to a complementizer. This particle is also used after verbs of knowing (Sridhar 1992:145). In Irish (Gaelic), direct and indirect question are identical both on the lexical and on the syntactic level (Ó Siadhail 1989: 321):

d')	<i>An</i>				<i>raibh</i>	<i>tú</i>
	<i>sásta?</i>					
	INTER(ROGATIVE PARTICLE)				bc-PAST	you
	content					
	‘Were you content?’					
d'')	<i>Chur</i>	<i>sé</i>	<i>ceist</i>	<i>ort</i>	<i>an</i>	<i>raibh</i>
	<i>sásta.</i>					<i>tú</i>
	put	he	question on-you	INTER.	bc-PAST	you
						content
	‘He asked you if you were content’					

Contrary to the substratist position, Denison (1998: 245-246) offers a superstratist explanation. He points out that embedded inversion existed already

in the superstrate, that is in 18th/19th century English under certain conditions even in literary style – as in example e).

e) *and Mr Casaubon had never himself seen fully what was the claim upon him. (Eliot, Middlemarch)*

According to Quirk et al. (1985: 1052) inversion occurs here, because the *wh*-work *what* functions as the subject complement. They add that in dialectal usage, it is more common and not restricted to certain conditions (*ibid.*).

Another reason for the presence of embedded inversion in contact varieties is that it might be a feature of the interlanguage of second-language learners in which the word order of direct questions is transferred to indirect questions. This is especially important in the acquisition of English as a second language by speakers whose native language does not make use of inversion in direct questions (see, e.g., Schmied 2004; Sand 2005 on New Englishes; Dulay & Burt 1974; Braidı 1999). Miller & Weinert (1998: 83) as well as McDavid & Card (1972: 105) therefore consider the occurrence of embedded inversion as the failure of the speaker to construct a correct indirect question.²

Another important factor is the fact that embedded inversion represents a blur between direct and indirect questions (see Biber et al. 1999: 920-921). Hence, Denison (1998: 246) calls clauses with embedded inversions “semi-quotations”. Examples of embedded inversion in the present tense may have the same overt structure as a reporting clause followed by direct speech – without directly addressing the hearer as illustrated in f):

f) *Yeah I'm asking is it okay (ICE-EA: cl-lessK)*
vs. *I'm asking, "Is it okay?"*

This makes embedded inversion very similar to free indirect speech as used as narrative device in fiction (see Fludernik 1993: 152-153) with the small difference that embedded inversion is not free, but subordinated to a matrix clause. Evidence for the fact that indirect and direct speech are not distinct categories, but may be blurred by speakers is also found in instances such as g):

g) *I was asking his wife how does she feels [sic!] now (ICE-IND: S1B-035)*

² Here we would like to anticipate from our analysis the sheer fact that over 90% of embedded clauses exhibit SV word order even in spontaneous speech which shows that all speakers included in the data base are able to construct indirect questions according to the Standard English pattern.

This speaker freely mixes features of indirect and direct speech. The direct question represented in g) is *How do you feel now?*. In example g) the speaker has shifted the deictic expression for person reference (for his/her interlocutor's wife) from *you* to *she*, but the sentence does not show sequence of tenses: although the reporting clause is in the past tense, the auxiliary *does* appears in the present tense. Thus, g) blurs direct and indirect speech by exhibiting both features of direct speech (word order, present tense) and features of indirect speech (person shift, intonation pattern) in one sentence.

The close relation of embedded inversion and direct questions or direct speech is supported by the fact that in British English, embedded inversion in general is more frequent with verbs that can function as the verb in a reporting clause for direct speech such as *ask* and *wonder* than with verbs that are at least very unusual to occur in this position such as *know* and *see* (Kolbe 2001: 56-74). Compare f) above with h)-i)

- h) *I'm wondering, "Is it okay?"*
- i) *?I don't know, "Is it okay?"*
- j) *?I came here to see, "Is it okay?"*

The hypothesis that there is an interconnection between reports of direct questions and the use embedded inversion also receives support by the claim that it is only possible in statements with question orientation, i.e., when the embedded interrogative clause actually expresses a clear lack of information (see Ohlander 1986: 972-973; Huddleston & Pullum 2002: 981). This would be displayed in all examples from our data: a), e), f) and g). Speakers in a) and f) lack the information whether something is okay, the subject in e) lacks the information about what is claimed while the speaker in g) lacks the information whether the wife of her interlocutor in the reported speech event is okay.

It is in these instances that speakers may also transfer direct question word order to indirect questions. If it is an overgeneralisation, however, one might wonder why the word order remains "direct" while deictic expressions (including tense) can be shifted.

However, examples such as k) show that embedded inversion is possible in clauses that do not express a lack of information, i.e. which are not indirect questions in that they do not render a formerly direct question or an underlying question, an urge to gather information previously not available (see Kolbe 2008: 150, 160-161).

k) *you know what am I like when it comes to writing letters (ICE-IRE WIB-004 Thailand letters 2)*

The lack of semantic question orientation in k) is created by the positive matrix clause *you know*. However, the embedded *wh*-clause *what am I like* in k) alone is a perfectly formulated Standard English direct question.

A formal feature that is often reported to influence the use of embedded inversion is the question type, *wh*-interrogatives vs. yes-no interrogatives (see, e.g. Hilbert 2008). However, while Platt, Weber and Ho (1984: 128) state that embedded inversion occurs only in *wh*-interrogatives, Henry (1992: 282) reports that a majority of her Belfast informants judge embedded inversion ungrammatical in *wh*-interrogatives but not in their yes-no equivalents.

In sum, the following aspects emerge as probable sources of embedded inversion:

i) It forms a blur between direct and indirect speech by retaining direct question word order on the one hand, but employing deictic shifts as in reported speech on the other hand.

ii) There is evidence that it is similar to structures in various substrate languages (Celtic and African languages); but it also has been attested in the superstrate language English even in more formal styles (fiction writing) in earlier centuries.

iii) The transfer of direct question word order to embedded clauses also appears to be a factor of L2 acquisition.

These factors are certainly not mutually exclusive. The blur between direct and indirect speech at the level of word order resembles a stylistic device and has apparently existed in English for centuries. The presence of identical word order in direct and indirect questions in substrate languages may have reinforced a structure in the superstrate just as much as an overgeneralisation in L2 acquisition by speakers whose native languages do not invert in direct questions.

3. Data

Our database consists of all available components of the International Corpus of English (ICE), i.e.

- Great Britain
- New Zealand
- Ireland (Northern Ireland and Republic of Ireland)

- Jamaica
- Singapore
- Hong Kong
- India
- Philippines
- East Africa

Each component contains 1 million words in 500 texts from various text types, comprising spoken and written data as well as public and private communication.

These are the text types represented in ICE:

- s1a: private dialogues (conversations and phone-calls)
- s1b: public dialogues (class lessons, broadcast discussions, ...)
- s2a: unscripted monologues (commentaries, legal presentations, ...)
- s2b: scripted monologues (broadcast news, non-broadcast talks, ...)
- w1a: non-printed written texts / student writing (essays, exam scripts)
- w1b: non-printed written texts / letters (social letters, business letters)
- w2a: printed academic texts (humanities, social sciences, ...)
- w2b: printed popular texts (humanities, social sciences, technology, ...)
- w2c: printed reportage (press reports)
- w2d: printed instructional texts (administrative writings, skills/hobbies)
- w2e: printed persuasive texts (editorials)
- w2f: printed creative texts (novels)

(see <http://www.ucl.ac.uk/english-usage/icc/design.htm>, last accessed 25 May 2009)

As the subcorpora of ICE have been compiled according to the same compilation guidelines, they are ideally suited to comparative analyses across varieties. It is important to point out, however, that previous research has mainly been conducted on the basis of anecdotal or dialectal data, while the ICE corpora represent more standard usage by educated speakers. The inclusion of rather informal written and spoken text types such as face-to-face conversations and private letters should compensate this difference. We thus expect to find more instances of embedded inversion in these text categories.

4. Methods

4.1. Variables

In order to calculate the determinants of embedded inversion, we identified factors that might possibly influence its use. These factors were then included as

independent variables in logistic regression analysis. Logistic regression is a probabilistic statistical analysis that allows to compare the strength of different variables and the degree to which they influence the odds of embedded inversion. Their strength is rendered by the odds ratio (OR).

The variables that are supposed to determine the use of embedded inversion are:

-QUESTION TYPE: There is (contradictory) evidence that with regard to embedded inversion yes/no indirect questions might act differently than *wh*-questions (Hilbert 2008, Henry 1995: 106, Platt, Weber & Ho 1984: 128).

-REGION: Filppula (2000) and Kolbe (2001) find that in the British Isles, embedded inversion is more frequent in the Celtic Englishes (Irish, Welsh and Scottish English) than in English English, which supports the hypothesis of substrate influence, confirmed by the lexical distribution of embedded inversion in Kolbe (2001).

- (matrix) VERB: Kolbe (2001) shows that embedded inversion in general is more frequent after the matrix verbs *ask* and *wonder*, which may also control direct questions more easily than *know* and *see*. This variable controls for differences between all verbs.

- Is the verb in the embedded clause BE? For the variable WHATS, which was excluded after preliminary analyses, we distinguished between clauses whose main verb is *be* and others. This variable proved to be a valuable factor in the final analyses.

- Hence also the variable SUBJECT LENGTH. One of the characteristics that accompanies embedded inversion even in Standard English is long subjects, e.g. in e) (Quirk et al. 1985: 1052 note). Longer subjects should favour inversion. The length is measured by the number of words which constitute the subject.

- Does the embedded clause contain *know/see what you/I MEAN*? We suppose that embedded inversion is highly unlikely in interrogative clauses which frequently complement discourse markers such as *(do) you know/see what I mean, I know/see what you mean*. The high frequency of these items and their discourse function should lead to a collocational restriction in which embedded inversion becomes nearly impossible.

- WHATS: *what's* represents a similar collocational restriction. It is used as one chunk rather than interrogative pronoun plus *is*, so that *I don't know what's the difference/the matter* is unlikely to be changed into *I don't know what the difference / matter is*.

- TEXT TYPE (ICE categories: s1a, s1b, s2a, s2b, w1a, w1b, etc.): Since embedded inversion is not considered grammatical in Standard English, we

hypothesise that more formal text types prevent its use. Hence, embedded inversion is expected to be less likely in scripted (s2b) than in unscripted (s2a) monologues, and in printed (w2) vs. non-printed material (w1).

- ORIENTATION (question vs. answer orientation): Although there are counterexamples to the claim that embedded inversion is possible only in embedded interrogative clauses with question orientation, these might be statistically irrelevant. As embedded inversion effects a blur between direct and indirect speech which is represented by embedded inversion, it might still have a significant influence on the use of embedded inversion whether the interrogative clause represents an actual or underlying question or not.

Previous findings suggest that certain groupings within individual variables might be important. Whether the variety is acquired as first or second language could lead to groups within REGION. *Know* and *see* are expected to act differently from *ask* and *wonder* and could represent different groups within VERB. Also, the ICE TEXT TYPES could be grouped into larger blocks (spoken vs. written, public vs. private, s1/s2/w1/w2). We created these different kinds of sub-variables, but they did not yield better results in any analysis than the original, more fine-grained variables.

We also included interactions between individual variables in the logistic regression models. An example for an interaction might be the fact that question orientation is a significant factor for the use of embedded inversion only if the embedded clause is controlled by *know*. This would represent an interaction between the variables VERB and QUESTION TYPE.

The independent variable is EI (embedded inversion): whether the clause contains embedded inversion or not and all independent variables are considered likely to favour or disfavour its use. Since it is only the absence or presence of embedded inversion that is captured in this variable, our logistic regression models are binary logistic regressions.

4.2. Compilation of the Database

In order to find possible loci of embedded inversion, concordance searches of the most frequent matrix verbs of indirect questions, *know*, *ask*, *wonder* and *see* (cf. Biber et al. 1999: 686) were conducted. From the concordances we extracted all embedded interrogative clauses manually.

If there was evidence against the embedded status of an interrogative clause, it was not included in the data base. Evidence against embedding were: i) lack of shifts for indirect speech; ii) punctuation: a comma, full stop or semicolon between matrix clause and embedded clause or a question mark following the embedded clause; iii) orthography: upper case letter in the first

word of the interrogative clause. If any of these features occurred together with subject-auxiliary inversion, the clause was removed from the database. In clauses without subject-auxiliary inversion, two additional features against embedding lead to the exclusion of the clause. This procedure is of course best suited to deal with written data, as the transcribers of the spoken texts may have been inconsistent in their treatment of embedded inversion, but we considered it best to follow a strict protocol rather than relying on intuitive judgments, as the original soundfiles or recording were not available to us. It is thus possible that the analysis is slightly biased in favour of the standard variant.

After preliminary analyses, we also excluded all clauses containing “what’s” from the database, because this lexical chunk does not allow variation. Without *what’s* clauses, the database consists of 6,590 clauses.

5. Analysis and Results

5.1 Analysis

In general, embedded inversion is the largely less frequent option in the database. Overall, 93.1% of the clauses have subject-word order and only 6.7% show embedded inversion. In 13 cases, we could not determine whether embedded inversion was present or not, usually because the subject or the verb was missing. An example is 1):

1) *Asking why isn't always such a bally great thing. young woman. he commented, focusing at a (ICE-Ireland, W2F-020)*

Since logistic regression analysis (here in SPSS) does not include missing cases, 6577 cases form the basis for the calculation of cases. The underlying “null hypothesis” for the following analysis is that embedded inversion is equally frequent across all text types, after all matrix verbs, in all varieties, independent of the question type etc.

As an inferential analysis, logistic regression determines which factors increase the likelihood of the occurrence of embedded inversions. Here, factors refer to the different values of the independent variables. The value “Hong Kong” in the variable REGION is the factor REGION(HongKong). On the one hand, the effect on likelihood is measured via the odds ratio (OR) of each independent variable. The higher the odds ratio of a variable, the more likely does it make embedded inversion, i.e. the higher are the odds of embedded inversion. There is no upper limit. Note that the odds ratio 4 of one factor as

opposed to the odds ratio 2 of another factor does not mean that the former is twice as strong. This is because an odds ratio of 1.0 means that the factor has no effect and odds ratios between 0 and 1 indicate a disfavoured effect on the use of embedded inversion. So whereas there is no upper limit, there is a lower limit and the range for disfavoured factors is much smaller than that for favouring factors (Pampel 2000:18-21)

The overall strength of the following logistic regression models is evaluated by three parameters:

- its overall statistical significance (all of the discussed models are significant at $p > 0.05$)
- the correlation coefficient “Nagelkerke R^2 ”. It ranges from 0 to 1; an R^2 measure of 1 would mean that all of the variability of the data is explained by the model. Values of around 0.2 are generally accepted as sufficient (cf. Pampel 2000: 53).
- whether the percentage of correctly predicted cases is increased. A good model should identify factors favouring or disfavoured embedded inversion in such a way that it allows to better predict whether a clause exhibits inversion or not. The baseline percentage is 93.2% – the percentage of clauses (out of 6,577) with standard uninverted word order. If one assumed that embedded inversion did not occur, that would mean a correct prediction in 93.2% cases. Since that is already a high percentage, all models increase the correctly predicted percentage of cases by only 0.1 to 93.3%.

The analysis also determines whether the influence of a factor is significant. Some factors may have high odds ratios, but not an actual effect. This is rendered by the absence of statistical significance.

5.2 Results

This section contains summaries of the two best logistic regression models for our data, one without interactions between variables (with only “main effects”) and the second including interactions. The model summaries present only statistically significant effects. These are calculated against a constant, which is automatically generated. Our constant is a yes-no interrogative clause (and thus naturally no occurrence of *know/see what you/I mean*³) from Singapore introduced by *wonder* whose verb is not a form of *be* and it occurs in a novel.

³ Although thus the variables QUESTION TYPE and MEAN partly exclude each other (when the TYPE is yes-no, MEAN cannot occur), these variables are not collinear (VIF 1.0) (cf. Menard 2002: 75-78).

Table 1 displays the significant main effects that increase the odds of embedded inversion are QUESTION TYPE(*wh*-clauses), BE(*be*), VERB(*ask*), REGION(Ireland) and TEXT TYPE(spoken, written and non-printed or press reports). Subject-auxiliary inversion in the corpus overall is more likely when the embedded clause is a *wh*-clause, when its main verb is *be*, when the matrix verb is *ask*, when the clause is from Ireland, and when it occurs in a spoken text, in a non-printed written text or in a press report. Nagelkerke R^2 of this model is 0.26.

These effects support hypotheses and findings from previous research. Subordinate Standard English yes-no interrogatives are introduced by *if* or *whether*. The fact that these have to be omitted for embedded inversion appears to inhibit the non-standard variant. Embedded inversion is more likely when its matrix verb could function as reporting verb of direct speech (even more if it is *ask* rather than *wonder*). Furthermore, less careful preparation of the utterance in spoken, unscripted and non-printed texts facilitates embedded inversions. It is not spoken language alone, in which processing efforts complicate keeping the distinction between indirect and direct speech. Embedded inversion also occurs in written language, when writers have time to consider their linguistic choices. Press reports thus appear less carefully prepared than novels, at least they are produced under considerably greater pressure of time than the latter. The analysis also provides evidence for substrate influence in Irish English. Ireland is the only region that increases the odds for embedded inversion and it is the only region with a Celtic language background in the data. (The possible influence of Scottish Gaelic and Welsh cannot be rendered in ICE-GB.) The strongest factors favouring embedded inversion is the text type, specifically the dialogues (s1a).

The significant main effects that reduce the odds of embedded inversion are MEAN(*what you mean*), VERB(*know*, *see*) and REGION(Great Britain, New Zealand). *Know/see what you mean* as discourse marker reassuring the speaker of their listeners' attention does not allow embedded inversion. The exceptions in the analysis are clauses that fit our criteria for the variable, but do not function as discourse markers:

m) *We have seen what do you mean by solvent, is <,> in which the substance gets dissolved* (ICE-India, S1b-004.txt)

Further, the matrix verbs *know* and *see* that usually do not introduce direct speech also disfavour the use of embedded inversion as well as the regions in the data that in general represent the most homogeneous speech communities with the least substrate influence, that is Great Britain and New Zealand. The strongest factor disfavouring embedded inversion is the matrix verb *know*.

Although this appears to be related to the fact that positive *know* does not allow question orientation, the semantic orientation of a sentence does not have a significant influence of the odds of embedded inversion.

Table 2 shows the significant factors in the second logistic regression model which also includes interaction effects. The respective Nagelkerke R² is 0.31, so the overall model is slightly better than the model with only main effects. Main effects that still increase the odds of embedded inversion again are QUESTION TYPE (*wh*-clauses), REGION(Ireland) and TEXT TYPES(spoken, non-printed written texts and press reports). These are thus the most stable predictors of embedded inversion.

Interactions that increase the likelihood of embedded inversion are BE(*be*) *SUBJECT LENGTH, BE(*be*)*VERB(*ask, know, see*) and REGION(Hong Kong)*VERB(*see*). Embedded inversion becomes more frequent with each additional word in the subject noun phrase in combination with the main verb *be*. Compared to *wonder*, all other matrix verbs increase the odds of embedded inversion when the main verb in the controlled clause is *be*. When an embedded interrogative clause in the data from Hong Kong is controlled by the matrix verb *see* it is more likely to show subject-auxiliary inversion. The strongest factor is again the text type, in the form of the conversations (s1a).

MEAN(*what you mean*) and VERB(*know, see*) are still factors that disfavour the use of embedded inversion. Interactions additionally reduce the probability of embedded inversion are interactions between the verb *be* with the regions Hong Kong, Ireland, Jamaica and Philippines (in contrast to Singapore), as well as interrogative clauses controlled by *see* in the data from New Zealand and Ireland and the matrix *know* in the data from Ireland (when compared to *wonder*). The strongest inhibiting factor is the verb *know*.

The interaction effects of REGION, BE and SUBJECT LENGTH qualify the main effects of Great Britain, New Zealand, *to be* and *ask* of the first regression model, though the differences on the overall quality of the model are minimal.

Ireland remains a stable factor favouring embedded inversion. However, embedded inversion is not more likely after *know* and *see*, but instead used less frequently. The regional preference therefore supports the assumption of Celtic substrate influence, but this is not shown by a more equal distribution across all matrix verbs. The higher use of embedded inversion after the matrix verbs *know* and *see* seems to be a peculiar feature of the English spoken in the Scottish Highlands and Hebrides (see Kolbe 2001: 58-68).

In sum, although embedded inversion is the less frequent word order option in embedded interrogative clauses in general, it occurs in all varieties under analysis. It is less frequent in those countries with comparatively little language contact, Great Britain and New Zealand, and it is especially strong in

Ireland. Thus it appears to be a feature of learner language in general. However, the Gaelic substrate in Ireland seems to have had more impact on this aspect of English grammar than the respective substrate languages of the New Englishes.

Subject-auxiliary inversion is most likely in spoken text types. However, it also occurs in formal and written texts. No text type has a significant inhibiting effect on embedded inversion. A blur between direct and indirect speech most easily occurs in spoken language, where matters of punctuation do not play a role. The only caveat are the transcription conventions applied in the compilation of the ICE corpora which could not be controlled for the purposes of the present study.

With regard to syntactic determinants, embedded inversion is more likely in *wh*-clauses, as long as they do not occur in *know/see what you mean*. The matrix verbs *know* and *see* which do not commonly function as reporting verbs of direct questions also control embedded inversion less frequently. This provides further evidence in support of the view that direct and indirect speech are not two distinct categories, but that they represent ends of a continuum, which also includes the narrative devices of literary free direct and free indirect speech. Inversion of verb and subject is also more likely when the main verb of the subordinate interrogative clause is *be* and no periphrastic *do*-support is necessary for the inversion especially with longer subjects. However, the positive influence of *be* on the likelihood of embedded inversion has significant regional restrictions.

6. Conclusions

What do our data say about the distribution and explanatory factors for embedded inversion? The fact that it is significantly more likely in less carefully planned texts and less likely after matrix verbs that do not typically introduce direct speech first of all provides evidence for the use of a blur between direct and indirect speech.

As embedded inversion is less likely in Great Britain and New Zealand (areas in which potential superstrate influence should show more clearly), it is more likely caused by the influence of substrate languages or at least presence of language contact. In varieties with a greater degree of language contact it is generally more frequent, which also indicates the importance of learner language in the emergence of this structure. The Gaelic substrate in Ireland appears to be a factor that further enhances the use of embedded inversion. As the inverted word order was already possible (at least in some contexts and structures) in the English superstrate, substrate influence may have led to a wider range of uses. It

thus becomes clear that any monocausal explanation or hypothesis is likely to fall short of fully accounting for the feature in question.

In addition to the analyses presented here, there is also an interesting and seemingly paradoxical interplay between the lack of inversion in direct questions and the occurrence of embedded inversion in the New Englishes (see Sand 2005: 188-197). This certainly merits further investigation, especially in the area of psycholinguistic research on individual speakers' mental grammar.

Our findings present only one small step in the description of grammatical variation across Englishes worldwide. However, it has become clear that the mere attestation of a certain feature is just one small part of a very complex picture. Only a more fine-grained statistical evaluation can help to identify the factors which promote or inhibit the use of a specific feature such as embedded inversion in a statistically significant way.

References

- Biber, Douglas et al. (eds.). 1999. *Longman Grammar of Spoken And Written English*. London: Longman.
- Braidi, Susan. M. 1999. *The Acquisition of Second-Language Syntax*. London: Arnold.
- Catford, J. C. 1957. The Linguistic Survey of Scotland. *Orbis* 6:105-121.
- Denison, David. 1998. 3. Syntax. In Suzanne Romaine (ed.), *The Cambridge History of the English Language, Vol. IV: 1776-1997*, 92-329. Cambridge: CUP.
- Dulay, Heidi & Marina Burt. 1974. Natural Sequences in Child Second Language Acquisition. *Language Learning* 24:253-278.
- Filppula, Markku. 2000. Inversion in Embedded Questions in Some Regional Varieties of English. In Ricardo Bermúdez-Otero et al. (eds.), *Generative Theory And Corpus Studies: A Dialogue from 10 ICEHL*, 439-453. Berlin: Mouton de Gruyter.
- Filppula, Markku. 2004. Irish English. Morphology and Syntax. In Bernd Kortmann et al. (eds.), *A Handbook of Varieties of English, Vol. 2: Morphology and Syntax*, 73-101. Berlin: Mouton de Gruyter.
- Fludernik, Monika. 1993. *The Fictions of Language and the Languages of Fiction: The Linguistic Representation of Speech and Consciousness*. London & New York: Routledge.
- Harris, John. 1993. The Grammar of Irish English. In Leslie Milroy & James Milroy (eds.), *Real English: The Grammar of English Dialects in the British Isles*, 139-186. London: Longman.
- Henry, Alison. 1995. *Belfast English And Standard English*, New York & Oxford: OUP.

- Hilbert, Michaela. 2008. Interrogative Inversion in Varieties of English: A Case of Angloversals? Paper presented at ISLE 1, Freiburg, 8 October 2008.
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: CUP.
- Kolbe, Daniela. 2001. Embedded Inversion in the North of the British Isles, unpubl. M.A. thesis, Albert-Ludwigs-Universität Freiburg.
- Kolbe, Daniela. 2008. Complement Clauses in British Englishes, unpubl. Ph.D thesis, Universität Trier.
- McDavid, Virginia Glenn & William Card. 1972. Problem Areas in Grammar. In A. L. Davis (ed.), *Culture, Class and Language Variety: A Resource Book for Teachers*, 89-132. Chicago: Center for American English. Illinois Institute of Technology.
- Menard, Scott. 2002. *Applied Logistic Regression Analysis*. 2nd ed. Thousand Oaks, London & New Delhi: SAGE.
- Miller, Jim. 1993. The Grammar of Scottish English. In James Milroy & Leslie Milroy (eds.), *Real English: The Grammar of English Dialects in the British Isles*, 99-138. London & New York: Longman.
- Miller, Jim & Regina Weinert. 1998. *Spontaneous Spoken Language: Syntax and Discourse*. Oxford: OUP.
- Ohlander, Sölve. 1986. Question-Orientation Versus Answer-Orientation in English Interrogative Clauses. In Dieter Kastovsky & Aleksander Szwedek (eds.), *Linguistics across Historical and Geographical Boundaries*, 963-982. Berlin & New York: Mouton de Gruyter.
- Pampel, Fred C. 2000. *Logistic Regression: A Primer*. Thousand Oaks, London & New Delhi: SAGE.
- Parry, David. 1999. *A Grammar and Glossary of the Conservative Anglo-Welsh Dialects of Rural Wales*. Sheffield: The National Centre for English Cultural Tradition.
- Penhallurick, Robert. 1991. *The Anglo-Welsh Dialects of North Wales*. Frankfurt am Main: Lang.
- Platt, John T., Heidi Weber & Mian Lian Ho. 1984. *The New Englishes*. London: Routledge.
- Quirk, Randolph et al. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Sabban, Annette. 1982. *Gälisch-Englischer Sprachkontakt: Zur Variabilität des Englischen im gälischsprachigen Gebiet Schottlands. Eine empirische Studie*. Heidelberg: Groos.
- Sand, Andrea. 1999. *Linguistic Variation in Jamaica: A Corpus-Based Study of Radio and Newspaper Usage*. Tübingen: Narr.
- Sand, Andrea. 2005. Angloversals? Morphosyntactic Parallels in Contact Varieties of English, unpubl. Habilitationsschrift, Albert-Ludwigs-Universität Freiburg.

- Schmid, Josef. 2004. East African English (Kenya, Uganda, Tanzania): Morphology and Syntax. In Bernd Kortmann et al. (eds.), *A Handbook of Varieties of English, Vol. 2: Morphology and Syntax*. 918-930. Berlin: Mouton de Gruyter.
- Schmid, Josef. 2008. East African English (Kenya, Uganda, Tanzania): Morphology and Syntax. In Rajend Meshtrie (ed.), *Varieties of English, Vol. 4: Africa, South and Southeast Asia*, 451-471. Berlin: Mouton de Gruyter.
- Simo Bobda, Augustin. 1998. *The Indigenization of English in Cameroon and New Englishisms*. Essen: LAUD.
- Sridhar, S. N. 1992. The Ecology of Bilingual Competence: Language Interaction in the Syntax of Indigenized Varieties of English. *World Englishes* 11:141-150.
- Sridhar, S. N. 1996. Toward a Syntax of South Asian English. In Robert J. Baumgardner (ed.), *South Asian English: Structure, Use and Users*, 55-69. Urbana, IL: U of Illinois P.
- Tay, Mary. 1982. The Uses, Users and Features of Singapore English. In John Pride (ed.), *New Englishes*, 51-70. Rowley, MA: Newbury House.
- Thomas, Alan R. 1997. The Welshness of Welsh English: A Survey Paper. In Hildegard L. C. Tristram (ed.), *The Celtic Englishes*. 55-85. Heidelberg: Winter.

Table 1: Significant main effects

Factor	OR
QUESTION TYPE(<i>wh</i> -clauses)	2.61***
BE(<i>be</i>)	2.95***
MEAN(<i>what you mean</i>)	0.18*
VERB(<i>ask</i>)	1.53**
VERB(<i>know</i>)	0.16***
VERB(<i>see</i>)	0.21***
REGION(Ireland)	1.46*
REGION(Great Britain)	0.30***
REGION(New Zealand)	0.30***
TEXT TYPE(s1a)	36.33***
TEXT TYPE(s1b)	34.82***
TEXT TYPE(s2a)	22.21***
TEXT TYPE(s2b)	12.49**
TEXT TYPE(w1a)	15.26**
TEXT TYPE(w1b)	11.44**
TEXT TYPE(w2c)	7.8*

* significant at $p < 0.05$

** significant at $p < 0.01$

*** significant at $p < 0.001$

Table 2: Significant main effects and interactions

Factor	OR
QUESTION TYPE(<i>wh</i> -clauses)	2.7***
MEAN(<i>what you mean</i>)	0.18*
VERB(<i>know</i>)	0.11***
VERB(<i>see</i>)	0.16**
REGION(Ireland)	7.9***
TEXT TYPE(s1a)	37.73***
TEXT TYPE(s1b)	35.18***
TEXT TYPE(s2a)	23.48***
TEXT TYPE(s2b)	12.58**
TEXT TYPE(w1a)	15.28**
TEXT TYPE(w1b)	10.32**
TEXT TYPE(w2c)	7.7*
BE(<i>be</i>)*SUBJECT LENGTH	1.03*
BE(<i>be</i>)*VERB(<i>ask</i>)	3.28***
BE(<i>be</i>)*VERB(<i>know</i>)	10.35***
BE(<i>be</i>)*VERB(<i>know</i>)	10.90***
BE(<i>be</i>)*REGION(Hong Kong)	0.43*
BE(<i>be</i>)*REGION(Ireland)	0.46*
BE(<i>be</i>)*REGION(Jamaica)	0.22*
BE(<i>be</i>)*REGION(Philippines)	0.35**
REGION(Hong Kong)*VERB(<i>see</i>)	15.87*
REGION(Ireland)*VERB(<i>see</i>)	0.16*
REGION(Ireland)*VERB(<i>know</i>)	0.09***
REGION(New Zealand)*VERB(<i>know</i>)	0.13*

* significant at $p < 0.05$