Structuring subjectivity in Asian Englishes
Multivariate approaches to mental predicates across genres and functional uses

Sandra C. Deshors
New Mexico State University

This study investigates the usage patterns of four near-synonymous mental predicates (*believe*, *guess*, *suppose* and *think*) across three Asian ESL (English as a Second Language) varieties as well as British and American Englishes. Using two multivariate techniques, multiple correspondence analysis and classification and regression tree analysis, the study shows the benefits of exploring cross-varietal variation through the lens of lexicalization patterns. The study also demonstrates that to make sense of semantic patterns it is crucial to account for extra-linguistic factors such as genre, as different ESL writers structure the meaning of *believe*, *guess*, *suppose* and *think* differently depending on their type of writing. Ultimately, in the broader context of the emancipation of ESL varieties, the results raise important questions about the developmental process of Asian Englishes and the place that semantic structure holds in this endeavor.

Keywords: Asian Englishes, classification and regression tree, mental predicates, multiple correspondence analysis, written genres

1. Introduction

As linguistic tools, epistemic stance expressions offer speakers the opportunity to convey their view of reality and to position themselves in relation to that particular reality. That is, through their uses of epistemic stance markers, speakers are able to offer insights into their personal view of a situation and they are able to indicate various degrees of commitment and attitudes towards their own statements (Kärkkäinen 2006) as well as express their commitment to the validity of their assertions (Lyons 1977; Aijmer 1997; Biber et al. 1999; Marín-Arrese 2009, 2011, 2015). Beyond their evaluative functions, however, epistemic markers also serve a...
social purpose. By presupposing systems of sociocultural values (Du Bois 2007), they are “a linguistically articulated form of social action whose meaning is to be construed within the broader scope of language, interaction, and sociocultural values” (Du Bois 2007: 139). As such, epistemic markers are interesting objects of study because their semantic configurations and uses may vary as a result of contexts of utterance and speakers’ cultural backgrounds.

Understanding the structure of the semantic notion of subjectivity and how the social aspect of stancetaking interacts with the construction of meaning has been of much interest to functional linguists over the past few years (e.g. Biber & Finegan 1989; Aijmer 1997; Hunston & Thompson 2000; Scheibman 2002; Kärkkäinen 2003a, 2003b; Englebretson 2007). Studying the phenomenon of stancetaking through the lens of socio-cognitive theoretical frameworks, scholars have been able to explore how speakers construe stance both as a social and as a cognitive phenomenon (e.g. Põldvere 2013). Importantly, this work has paved the way for recent empirical studies such as Krawczak (2014) in the area of dialectology. Adopting a sophisticated corpus-based perspective, Krawczak (2014) explores stancetaking across native English varieties, contrasting the British (BrE) and American (AmE) varieties with a view to understanding how mental predicates such as the near-synonyms believe, guess, suppose and think are construed in use, what functional components characterize their individual usage profiles and to what extent variation in their usage patterns is observed across speaker populations exposed to different socio-cultural contexts. Initially, Krawczak’s (2014) study stems from the observation that by selecting one of a number of synonyms of I think, a speaker can change significantly the character of a speech event and the expressive strength of a statement. With that in mind, exploring speakers’ choices of such predicates across English varieties is no trivial task as it can help us identify patterns of “domestication of meaning” (Okunrinmeta 2013) by different populations of English users.

In this context, the present study builds on Krawczak’s (2014) analysis in three different ways. Firstly, it expands her approach to believe, guess, suppose and think beyond native English varieties so as to include three varieties of institutionalized second-language Englishes (ESL) – namely, Hong Kong, Indian and Singaporean Englishes – and thereby compares Englishes with greater contrasts in their socio-cultural contexts. Secondly, following González et al.’s (2014) view that stancetaking patterns differ across speech and writing, I apply Krawczak’s (2014) approach (which approximates the spoken mode) to written data.1 Thirdly, while studies on stancetaking in the written register have primarily focused on the

1. In Krawczak (2014), the data used are blog-based. Therefore, they may be understood as approximating the spoken mode, by being quite interactive, natural and spontaneous in nature.
journalistic genre, the present study broadens the range of investigated types of writing to include correspondence, editorials, creative, academic, student, popular and instructional writing.

Methodologically, this study follows Krawczak’s (2014) footsteps in her application of multivariate methods to make sense of semantic structure and its variation across English dialects. The present work therefore adopts a two-step approach consisting of a multiple correspondence analysis and a classification and regression tree analysis. Although, over the past few years, this type of approach has increasingly been used in comparative corpus-based studies of ESL varieties, those studies mainly focus on exploring syntactic alternation patterns (e.g. dative alternation, gerund vs. infinitive verb complementation patterns). The present study is therefore the first of its kind to apply those methodologies to explore semantic structure across Asian Englishes. In what follows, I start by briefly presenting the conceptual make-up of the mental predicates believe, guess, suppose and think (Section 2.1). Secondly, I discuss the notion of semantic variation in the context of stancetaking (Section 2.2). Thirdly, I discuss the relevance of exploring semantic variation across Asian Englishes (Section 2.3). In Section 3, I present the methodological set-up of the study and in Sections 4 and 5 I present and discuss, respectively, the results of the analyses.

2. Adopting a semantic perspective on cross-varietal variation in ESL: The case of mental predicates

2.1 The conceptual make-up of epistemic stance expressions: Believe, guess, suppose and think

Broadly speaking, mental predicates pertain to the semantic domain of epistemic modality which is concerned with knowledge and belief (Kärkkäinen 2003a, 2003b) and it involves speakers making judgements about the truth of given propositions (Palmer 1990). Epistemic modality also involves linguistic elements expressing speakers’ attitudes to knowledge (Chafe 1986), such as (i) speakers’ assumptions or assessment (see Goodwin & Goodwin 1992; Goodwin 2006), (ii) speakers’ evaluation of possibilities (Lyons 1977; Nuyts 2001; see Labov & Waletzky 1967; Linde 1997; Conrad & Biber 2000; Hunston & Sinclair 2000; Thompson & Hunston 2000; Macken-Horarik & Martin 2003) and, in most cases, (iii) speakers’ confidence (or lack thereof) in the truth of the proposition expressed (Coates 1983). With regard to (iii), Sanders & Spooren (1996) note that epistemic markers differ with respect to their degree of certainty and subjectivity. For instance, for Krawczak (2014: 309), think and believe “make overt reference to the conceptualizer [i.e. speaker], being
subjective and semi-certain” and they are “more likely to be used in contexts where the speaker’s standpoint is strong, fully profiled and intrinsic to the subject”. However, with believe, although the predicate conveys a subjective attitude and some degree of supporting evidence towards the speaker’s statement, for Aijmer (1997: 17) “the evidence is incomplete or non-specific”. In contrast, guess and suppose have been shown to express a weaker view of reality and “they are likely to be used with statements whose epistemic status is difficult to judge” (Krawczak 2014: 309). In addition, mental predicates have been shown to differ with regard to reliability in the context of speakers’ knowledge. According to Aijmer (1997), mental predicates convey different degrees of reliability of knowledge and she shows that on a scale from stronger to weaker, the predicates can be organized as follows: believe, think, suppose and guess (see also Fortescue 2001). In this context, “replacing [think] with another near-synonymous item will change considerably the meaning of what is said” (Krawczak 2014: 305).

At a more conceptual level, the literature anchored in cognitive linguistics has shown that two concepts characterize stancetaking, namely subjectivity and negotiability. With subjectivity, the notion is to be understood in its sharp contrast with its counterpart, objectivity, and this, in relation to a “vantage point” (Langacker 1999), namely the perspective through which the conceptualizer views a given situation. When construed objectively, situations are conceived as ‘a focused object’; as such, those elements are considered “onstage” (Langacker 1999: 297). In contrast, situations construed subjectively are “offstage” or “implicit” (Langacker 1999: 297). With negotiability, personal opinions are not assumed to be set in stone; instead, they are assumed to be subject to negotiation between interlocutors. In this light, “a negotiable proposition […] involves an epistemic assessment – on the part of the virtual conceptualizer – concerning the existence […] of the profiled relationship” (Langacker 2009: 234).

2.2 Variation in the use of mental predicates: What are the triggering factors?

Hsieh (2009) has argued that the best balance in human communication is achieved when the form of language interacts with the function of discourse. In the process of achieving this balance, speakers can develop variation patterns when using linguistic items while adjusting to particular contexts of utterance. This is observed in Simon-Vandenbergen (1997), where several factors such as position, intonation and type of proposition are reported to trigger the use of I think with different functions in different contexts. Similarly, Aijmer (1997) shows that as a response to the demands of planning and conversational interaction, I think has
developed a number of new functions. Generally, epistemic stance markers are particularly sensitive to their register (writing vs. speech), genre and language (variety) of occurrence.

In terms of register, Gonzáles et al. (2014) demonstrate how function correlates with speech and/or writing as they observe that low certainty epistemic markers are more frequent in oral reports compared to written reports, indicating that when conveying their opinion and getting their points across, writers are more assertive than speakers. Further, the authors find that “both speakers and writers make use of different sources of knowledge depending on their epistemic stance […] thus adjusting and making use of different legitimizing strategies which are partially dependent on register” (Gonzáles et al. 2014: 18).

With respect to genre, Hsieh (2009) demonstrates how different genres and their different functional requirements lead to different uses of epistemic stance markers. For instance, focusing on the motivation for epistemic use in the discourse context of Chinese newspaper reporting, Hsieh (2009) shows how “epistemic marking fulfills different forces of stance taking in the journalist language” (Hsieh 2009: 28) and how ideology and standards observed in the journalistic society influence epistemic usage patterns. This leads Hsieh (2009: 28) to conclude that “by bringing to light the sensitivity of epistemic patterns to contextual factors, one may capture part of the homogeneous and heterogeneous qualities embedded in epistemic modal markers”. In the same spirit, Marín-Arrese (2015) also focuses on epistemic stance in journalistic discourse but she takes a cross-linguistic approach centered on English and Spanish. Across journalistic sub-genres (e.g. news reports, opinion columns), she discovers both similarities and differences in English and Spanish. For instance, in both languages, she finds epistemic stance markers to yield similar patterns of distribution in opinion and leading articles. However, mental predicates are an exception to this pattern, being more frequent in opinion columns and less frequent in news reports. For Marín-Arrese (2015), writers’ cultural differences in discourse practices may explain variation in epistemic usage patterns. Drawing on Bucholtz & Hall (2005), she tentatively claims that “particular stance-taking patterns may characterize cultural identities in discourse” (Marín-Arrese 2015: 224). Ultimately, Hsieh’s (2009) and Marín-Arrese’s (2015) studies raise the question to what extent the journalistic genre differs from other types of writing and whether those genres reflect speakers’ cultural backgrounds differently.

With regard to language varieties, modal expressions have been observed to vary at different linguistic levels and in relation to semantic and syntactic properties, pragmatic conditions on actual use, historical stages of grammaticalization and interplay with other grammatical categories (Chafe 1986; DeCarrico 1986; Bybee et al. 1994; Hsieh 2009). Focusing on native English varieties, Krawczak
(2014) documents cross-varietal variation in the use of mental predicates in British and American Englishes. For instance, she shows how, in BrE, believe is dissociated from non-argumentative statements expressing opinions and how it is associated with propositional statements whose epistemic content can be verified. In contrast, in AmE, “the predicate is predicated by positive evaluation, negotiability of the proposed state of affairs and statements expressing propositions that concern the real world” (Krawczak 2014: 319). As for think, the uses of the marker also vary across BrE and AmE. In BrE, it relates to emphatic and negotiable epistemic contexts where evaluations (rather than propositions) are expressed and where the conceptualized situation involves the future or an imaginary world. In AmE, think is more often associated with non-argumentative statements expressing personal opinions; although these statements are negotiable and objectively construed, they are not verifiable (Krawczak 2014).

2.3 Towards an exploration of semantic variation across Asian Englishes

Given the above-described context, contrasting native and second-language (ESL) varieties of English comes as the natural next step in understanding how different populations of English users structure and ultimately lexicalize the subjective notion of belief. Recently, corpus-based analyses contrasting ESL varieties have shown that differences between varieties of English manifest themselves in different degrees and at different levels of analysis (Bernaisch et al. 2014) and much of contemporary work in ESL has so far mainly focused on syntactic or lexicogrammatical features (Biermeier 2008; see Mukherjee & Schilk 2008; Mukherjee & Gries 2009; Schilk et al. 2012; Bernaisch et al. 2014; Bernaisch & Gries 2016 among others for studies exploring the lexicogrammatical interface). To date, ESL varieties have rarely been contrasted empirically from a semantic perspective (however, see Werner & Mukherjee 2012 for a corpus-based study on highly polysemous verbs in ESL); and although there are ESL studies focused on lexical phenomena (e.g. Biermeier 2009, 2014; Callies 2016), those tend to take a morphological perspective and focus on word-formation across ESL varieties. Lexico-semantic variation across ESL varieties is an area that has yet to be explored. Recent work by Okunrinmeta (2014) on Nigerian English (NigE) demonstrates the potential of this area of research. Providing an extensive list of lexicosemantic variations in NigE

---

2. I acknowledge Gablasova et al. (2015), who investigate epistemic stance in spoken L2 production. However, I do not discuss the study further because it is based on EFL varieties (i.e. foreign varieties of English spoken in countries such as France or Germany) rather than ESL varieties (i.e. indigenized varieties of English spoken in countries such as Singapore or Hong Kong where English is institutionalized).
showing how cultural social practices infiltrate the lexis of NigE, Okunrinmeta (2014) raises the question whether the semantic structure of lexical items can vary across different ESL varieties when those items are not rooted in concrete cultural practices but rather reflect the conceptualization of a subjective notion such as belief. The present study aims to address this question by investigating the usage patterns of believe, guess, suppose and think across three Asian English varieties as well as British and American Englishes.

Singaporean (SingE), Indian (IndE) and Hong Kong (HKE) Englishes are particularly well suited for this question, mainly because, based on Schneider’s (2007) dynamic-evolutionary model, those Asian English varieties are at different stages of evolution (or emancipation) in their developmental process towards nativization. Specifically, Schneider’s (2007) model assumes that the formation of postcolonial Englishes is driven by an underlying evolutionary process shared by all ESL varieties. This process operates whenever a language is transplanted and explains many similarities across ESL varieties. Importantly, “this process is characterized by a sequence of five characteristic stages that are associated with linguistic changes as well as the gradual emergence of locally characteristic linguistic patterns” (Deshors & Gries 2016: 213). According to the model, SingE, IndE and HKE differ in that they currently represent different stages of evolution, SingE being at a most advanced stage of evolution compared to IndE and HKE (phase IV; i.e. as the input language may have been retained as a (co-)official language and is used for intra-national contexts, the variety is undergoing a stabilization process), IndE being more advanced than HKE and less advanced than SingE (transitioning between phases III and IV; i.e. as the input language is becoming part of the local linguistic repertoire and the number of competent bilingual L2 speakers increases, the new English variety is beginning to develop accepted local standards), and HKE being the least advanced of the three varieties in question (transitioning between phases II and III; i.e. while the input language still determines language standards and norms in the variety, it is nonetheless becoming an integral part of the local linguistic repertoire) (Mukherjee & Gries 2009). In the context of the current work, investigating SingE, IndE and HKE together provides the opportunity to explore to what extent semantic structure correlates with ESL’s developmental stages. In light of all the above, the present study is set up to address the following research goals:

– to what extent the patterns of use of the mental predicates believe, guess, suppose and think differ across native and Asian Englishes;
– to what extent those different patterns also vary across written genres; and
– to what extent native and ESL writers (and Asian English writers specifically) conceptualize mental predicates differently across those genres.
3. Methodology

The methodology employed in the present study combines qualitative with quantitative analysis in the tradition known as the Behavioral Profile (BP) approach (Gries 2006; Divjak & Gries 2009; Divjak 2010; Glynn 2014a; Krawczak 2014). Broadly, this approach combines the statistical methods of contemporary quantitative corpus linguistics with a cognitive-linguistic and psycholinguistic orientation and is based on the idea that linguistic items can be characterized on the basis of their co-occurrence with other linguistic components. Ultimately, the profiling process involved in the method requires analysts to take into account a wide variety of semantic (and morpho-syntactic) features that all together contribute to the profile of a given linguistic item. For the purpose of this study, I used data from the International Corpus of English (ICE; see Greenbaum 1991). In Section 3.1, I explain how the data were extracted and annotated, and in Section 3.2, I present the statistical techniques employed in the present study.

3.1 Corpus data and annotation

All occurrences of believe, guess, suppose and think immediately preceded by a first person pronoun, a total of 908 contextualized occurrences, were extracted from the British, American, Hong Kong, Indian and Singaporean written subsections of ICE. The data were retrieved from all written genres represented in the corpus. However, for the purpose of the analysis, those different genres were conflated into the following seven categories: academic (including: academic humanities, academic social sciences, academic natural sciences and academic technology; files W2A-001 to W2A-040), correspondence (including: social correspondence and business correspondence; files W1B-001 to W1B-030), creative writing (files W2F-001 to W2F-020), instructional (including: administrative and skills/hobbies; files W2D-001 to W2D-020), popular (including: popular humanities, popular social sciences, popular natural sciences and popular technology; files W2B-001 to W2B-040), editorials (including: press news reports and editorials; files W2C-001 to W2C-020 and W2E-001 to W2E-010) and student writing (including: student untimed essays and student exams; files W1A-001 to W1A-020). The data were extracted using the software R (cf. R development Core Team 2015). Each extracted occurrence was checked manually for semantic relevance and occurrences that did not convey stance-taking were filtered out (e.g. When I think about Eric I have something to remember; I think of you often). Table 1 presents an overview of the distribution of mental predicates across sub-corpora after semantic check.
Table 1  Overview of the distribution of mental predicates across sub-corpora

<table>
<thead>
<tr>
<th></th>
<th>AmE</th>
<th>BrE</th>
<th>HKE</th>
<th>IndE</th>
<th>SingE</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>believe</td>
<td>26</td>
<td>26</td>
<td>20</td>
<td>4</td>
<td>19</td>
<td>95</td>
</tr>
<tr>
<td>guess</td>
<td>35</td>
<td>8</td>
<td>13</td>
<td>4</td>
<td>16</td>
<td>76</td>
</tr>
<tr>
<td>suppose</td>
<td>6</td>
<td>20</td>
<td>3</td>
<td>4</td>
<td>16</td>
<td>49</td>
</tr>
<tr>
<td>think</td>
<td>121</td>
<td>84</td>
<td>142</td>
<td>21</td>
<td>64</td>
<td>432</td>
</tr>
<tr>
<td>total</td>
<td>188</td>
<td>138</td>
<td>178</td>
<td>33</td>
<td>115</td>
<td>652</td>
</tr>
</tbody>
</table>

Each contextualized occurrence of the mental predicates was annotated for ten variables providing information on dialect, genre and functional contexts of use. These variables are listed in Table 2.

Table 2  Overview of the variables included in the study

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lemma</td>
<td>believe, guess, suppose, think</td>
</tr>
<tr>
<td>Variety</td>
<td>GB, HK, IND, SING, USA</td>
</tr>
<tr>
<td>Genre</td>
<td>academic, correspondence, creative, instructional, popular, editorials, student</td>
</tr>
<tr>
<td>EpistemicType</td>
<td>conviction, intention, opinion, prediction</td>
</tr>
<tr>
<td>EpistemicClass</td>
<td>evaluation, proposition</td>
</tr>
<tr>
<td>EpistemicMode</td>
<td>irreal, real</td>
</tr>
<tr>
<td>Argumentativity</td>
<td>argumentative, non-argumentative</td>
</tr>
<tr>
<td>Verifiability</td>
<td>non-verifiable, verifiable</td>
</tr>
<tr>
<td>Evaluation</td>
<td>negative, neutral, positive</td>
</tr>
<tr>
<td>Negotiability</td>
<td>negotiable, non-negotiable</td>
</tr>
</tbody>
</table>

The annotation scheme presented in Table 2 is borrowed from Krawczak (2014), which explores epistemic stance predicates in BrE and AmE. Krawczak (2014) developed the above scheme to explore subjectivity comprehensively through the lens of seven semantic variables, namely epistemic type, epistemic class, epistemic mode, argumentativity, verifiability, evaluation and negotiability (see Krawczak (2014) for a detailed description of these variables). Although I adopt Krawczak’s (2014) annotation scheme to ensure results comparability, it should be noted that, in itself, the scheme is not without any limitations. For instance, it does not account for the distinction between the use of mental state predicates as formulaic stance markers (where the main discourse contribution lies in the complement clause) and cases where the complement-taking predicate clause is the main point of the utterance.

With regard to epistemic type and epistemic class, both variables broadly refer to the notion of ‘epistemic’ in terms of stance types (i.e. evaluative, epistemic, intentional, etc.) and have to do with the predicate in the utterance and the function
that this predicate performs beyond conveying modality. In the case of epistemic type, the variable encodes whether the utterance conveys the speaker’s opinion, conviction, intention or whether s/he is making a prediction (as illustrated in examples (1) to (4)). Importantly, while, according to Aijmer (1997), the notion of epistemic type includes various senses and cases of extended meanings such as ‘belief’, ‘opinion’ and ‘subjective evaluation’, Krawczak (2014: 323) notes that, in her annotation scheme, ‘convictions’ and ‘opinions’ correspond to Aijmer’s (1997) notion of ‘belief’. However, it remains unclear whether Aijmer’s (1997) notions of ‘opinion’ and ‘subjective evaluation’ are actually accounted for in Krawczak’s (2014) scheme under the label ‘opinion’. For the purpose of the present study, I make no distinction between ‘belief’ and ‘opinion’. In this context, (1) is coded as an opinion because the information expressed pertains only to the speaker and does not so much concern the external world, as is the case in (2). Further, convictions involve a higher degree of commitment on the part of the speaker, as can be observed in (2), where the inclusion of the auxiliary do in I do believe strengthens the speaker’s belief. With regard to intention in (3) and prediction in (4), both semantic features have in common their orientation towards the future. However, they differ in that, contrary to prediction, intention is anchored in volition. More concretely, in (3), the speaker is willing to stop here whereas in (4), the fact that today is going to be a scorcher does not depend on the speaker’s intention.

(1) I think I was still tired because I had been sick
(epistemic type, opinion, usa:w1b-010)3

(2) I do believe that life is a continuous process of learning
(epistemic type, conviction, ind:w1b-014)

(3) Well, I think I will stop here, mail me soon
(epistemic type, intention, hk:w1b-002)

(4) I think today is going to be a scorcher
(epistemic type, prediction, gb:w1b-005)

With epistemic class, however, the focus is on whether the predicate contains an evaluation on the part of the speaker or a proposition (as illustrated in (5) and (6)).4 In (5), the predicate is coded as an evaluation as the subject assesses the chances that

3. Corpus references such as ‘usa:w1b-010’ include the following information: English variety (American English, in the current example), file number (w1b) and text number (010).

4. The categories in Krawczak’s (2014) scheme apply to the functional context of use of the mental predicates, that is, to features associated with the proposition expressed by the complement clause in complement-clause-taking constructions. This can be observed in (3), where the meaning of intention is invoked by the modal of volition in the complement clause.
a particular event (i.e. that) may occur and s/he comes to the conclusion that those chances are slim. In contrast, in (6), the predicate is coded as a proposition as the predicate lacks the evaluative characteristic found in (5) and the subject makes no assessment with regard to whether or not all will be attending the grand wedding.

(5) I suppose the chances of that are pretty darn slim
   (epistemic class, evaluation, usa:w2f-008)
(6) I guess they would all be attending the wedding
   (epistemic class, proposition, ind:w1b-006)

Moving on to epistemic mode, the variable has to do with the nature of the situation referred to in the utterance. Specifically, the variable encodes whether that situation is anchored in reality (i.e. real) as in Example (7) or whether the speaker locates it in the future or in some imaginary world (i.e. irreal) as in Example (8).

(7) I suppose that you are much too busy to do that
   (epistemic mode, real, sin:w1b-014)
(8) I suppose I could find out if I really wanted to
   (epistemic mode, irreal, usa:w1b-004)

The variables verifiability and argumentativity have to do with the type of knowledge that the speaker and interlocutor have of a particular event; that is, whether or not the statement made in the utterance can be verified (verifiable vs. non-verifiable), as illustrated in (9) and (10), or even challenged (argumentative vs. non-argumentative), as illustrated in (11) and (12).

(9) I suppose she used someone else’s address and phone number for the RSVP bit
    (verifiability, verifiable, gb:w2f-019)
(10) I suppose she meant his Dutch-tinted English
     (verifiability, non-verifiable, sin:w2b-007)
(11) I guess she’s happy, although poor
     (argumentativity, argumentative, usa:w1b-006)
(12) I think that the summer program is really well set up
     (argumentativity, non-argumentative, usa:w1b-003)

In (9), the predicated statement is verifiable in the sense that it is possible for anyone to check whether she used someone else’s contact details or not, for instance by going through records or databases. In contrast, in (10), the speaker does not have direct access to her state of mind and checking whether she meant his Dutch-tinted English or not is not something that can concretely be done. Argumentativity encodes whether or not the content of the proposition expressed
by the complement clause can be challenged by the interlocutor. With that in mind, (11) is coded as argumentative as whether she is happy or not is a state of affairs open for negotiation. For instance, the interlocutor may know her under particular circumstances that allow him/her to know how unhappy she is. In contrast, (12) is coded as non-argumentative as the proposition expresses a personal (i.e. entirely subjective) evaluative opinion that cannot be questioned.

With evaluation, the predicated event is assessed in terms of positivity, negativity or neutrality, as illustrated in (13, 14) and (15). In (13), the predicated event (i.e. sending and receiving letters that travel great distances) is presented positively by the speaker who qualifies the event as terrific. In contrast, in (14), the speaker’s negative evaluation of the event transpires through the use of the term decadent. Finally, in (15), the predicated event (i.e. the change in their underlying attitude towards nature) is presented as a fact and is not tinted by the speaker’s positive or negative assessment of the situation.

(13) I think it’s terrific to send and receive letters that travel great distances (evaluation, positive, usa:w1b-015)
(14) I guess it is decadent on my salary (evaluation, negative, usa:w2b-017)
(15) I don’t think their underlying attitude towards nature has changed (evaluation, neutral, hk:w2b-024)

Finally, let us turn to the last semantic variable, negotiability. This variable focuses on the question whether the statement made in the utterance leaves room for negotiation (i.e. for the addressee to challenge the statement; negotiable) or not (i.e. non-negotiable). While, in Krawczak (2014), the distinction between the notions of argumentativity and negotiability remains to be clarified, Langacker (2009) discusses negotiability in relation to “an interactive system” which comprises two main dimensions, polarity and illocutionary force. Broadly, this system is based on “the validity of the proposition expressed by the basic clause” (Langacker 2009: 232). In the context of the current study, the distinction between the negotiable and non-negotiable levels is illustrated in (16) and (17).

(16) I think you will forgive me, OK! (negotiability, negotiable, hk:w1b-004)
(17) I guess this is something I have to accept (negotiability, non-negotiable, sin:w2b-013)

While in (16) the addressee (i.e. you) is not obliged to forgive and is in a position to challenge the speaker’s belief, in (17) the use of the quasi-modal have to indicates that negotiating whether this is acceptable or not is simply not an option.
3.2 Statistical evaluation

Similarly to Krawczak (2014), the present study offers a two-step methodology that first involves the use of an exploratory technique, namely Multiple Correspondence Analysis (MCA), followed by the use of a confirmatory technique, namely Classification And Regression Tree (CART) analysis. However, the second step of the present analysis contrasts with Krawczak’s (2014) approach: while she adopts a regression approach as her chosen confirmatory technique, the CART technique used in the present study has been reported to handle low frequencies more efficiently than logistic regression approaches (Phelps & Merkle 2008). Other benefits of CART over regression are listed further below. In what follows, I present in more detail the two techniques used in the current study. I begin with MCA, followed by CART.

3.2.1 (Multiple) Correspondence Analysis

In the first step of the study, I performed an MCA to explore how different linguistic dimensions (e.g. genre, dialect, epistemic type, argumentativity) interact with one another when mental predicates are being used. MCA is an exploratory technique that follows a bottom-up approach and reveals frequency-based associations in complex corpus data (Glynn 2014a). Put more technically, using a distance-based clustering technique, MCA allows for the graphic representation of the structure of cross-tabulations on a multidimensional plane (Benzécri 1973, 1984; Greenacre 2007; Glynn 2014a). The advantages of this approach are two-fold. First, the technique helps identify associations between data points by way of visualization using biplots. These biplots depict degrees of correlation and variation based on the relative proximity of data points. The second advantage of this approach is specific to the ‘multiple’ type of Correspondence Analysis (as opposed to binary Correspondence Analysis). In contrast to ‘binary’ Correspondence Analysis, MCA allows researchers to assess the correlation between more than two factors simultaneously, therefore allowing them to identify (more) complex patterns potentially at play behind ESL users’ choice of mental predicates.

The MCA was conducted using the ca package for R (cf. R Development Core Team 2015). While the ca package offers three different types of MCA, namely indicator, Burt and joint, the joint option was selected as it has been shown to be superior both in terms of its ability to explain inertia and in the accuracy of the visualization (Greenacre 2006, 2007). In Correspondence Analysis, ‘inertia’ refers to the degree of variation in the data and it measures its total variance. Broadly

---

5. See Brezina (2013) for a study on certainty and uncertainty in spoken language with a focus on epistemic sociolect and idiolect and using a correspondence analysis approach.
speaking, inertia is calculated on the basis of observed and expected frequencies of co-occurrence; the higher the inertia, the better it explains the total variance in the data. As noted in Glynn (2014a: 450), the joint method “works by restricting the analysis to the cross-tabulations that typically contain the correlations of interest, those that explain the inertia”. At this point, it is important to bear in mind that despite its undeniable advantage to account for correlations across three or more factors at the same time, the method presents one main drawback, namely that these scores output by the analysis are often hard to interpret, mainly due to the fact that they “seriously underestimate the amount of accurately described variation, giving unnecessarily ‘bad’ results” (Glynn 2014a: 446; my emphasis). However, by combining MCA and CART, I anticipate to be able to unveil fine-grained patterns of use of mental predicates that so far have remained obscured in traditional frequency-based analysis.

3.2.2 Classification And Regression Tree
While MCA reveals potential associations between objects of a single variable, as a confirmatory method, CART identifies factors that lead to an outcome or a prediction (Glynn 2014b) and allows analysts to assess statistically to what degree the associations identified with the MCA are really present in the data. Based on Timofeev (2004) and Phelps & Merkle (2008), adopting the CART technique presents a number of advantages: (i) CART can handle small datasets (as previously mentioned), (ii) as a nonparametric statistical approach, CART only comes with a limited number of statistical assumptions, (iii) the method can be applied to various data structures and is exceptionally efficient in handling categorical predictors, (iv) the method involves automatic handling of stepwise variable selection, complexity reduction and interaction, and the CART algorithm itself identifies the most significant variables and discards non-significant ones, (v) the output provides rankings of variable importance, (vi) unlike other methods such as Principal Component Analysis or linear regression, CART easily handles outliers and (vii) the output of a CART analysis can easily be understood and interpreted. Despite these advantages, it should be noted that the CART method also comes with the drawback that even though tree models are generally robust, they can also be relatively unstable (i.e. small modifications of learning samples, such as the removal of several observations, can cause the tree to change radically). In the present study, the classification tree was computed using the *ctree* function in the party R package. The classification method was fed the model formula represented in (18), in which the variable Lemma is predicted on the basis of all the other above-described predictors (see also Table 2).
4. Results

As anticipated, the results of the MCA and CART analyses unveil usage patterns of mental predicates that differ widely across English varieties and written genres. As I show in the following subsections, the combination of the two statistical techniques with the qualitative analysis based on the annotation scheme developed in Krawczak (2014) makes it possible to draw a relatively nuanced picture of the uses of mental predicates in Asian and native Englishes and to observe the effects of different genres on the uses of the four mental predicates in focus.

4.1 Joint Multiple Correspondence Analysis

The output of the MCA is presented graphically in Figures 1 and 2. Note that both figures show the same output and that the latter is a close-up version of the former. As such, Figure 2 allows for greater visibility of the main cluster of factors in the upper part of Figure 1. More specifically, the two figures represent graphically the interactions between the mental predicates in relation to the dialect they occur in, written genre and all the semantic variables presented above. The plots should be interpreted by considering the relative distance between data points: the closer these points are to one another, the stronger the correlation between them. However, given that the plots flatten multidimensional distances onto a two-dimensional space (Glynn 2014a), the results of this analysis remain tentative and exploratory in nature.

Generally, the patterns presented in Figures 1 and 2 provide a promising starting point in the effort to explain how variety and genre influence writers’ lexical choices and what bearing these two factors have on the meaning of mental predicates. With an explained inertia of 70.3%, the patterns observed in the data can be approached with confidence (see Glynn 2014a for an in-depth discussion on acceptable inertia in MCA). Overall, it emerges that when all factors are considered simultaneously, the four lemmas in focus yield contrasting patterns of use across the English varieties. This is based on the facts that, in Figures 1 and 2, (i) the four mental predicates spread across different quadrants of the biplots, indicating that the predicates are different enough not to be clustered together (believe and think clustered together in the top right quadrant vs. suppose and guess clustered...
1. a first cluster of two lemmas, *believe* and *think*, which is located in the top right quadrant and associates most strongly with the BrE native variety;
2. a second cluster of two other lemmas, *guess* and *suppose*, which is located in the bottom left quadrant and associates most strongly with the AmE native variety;
3. a resemblance between SingE and AmE in their uses of *suppose* (as both varieties are plotted in the same quadrant and relatively close to the lemma in question);
4. a tendency for linguistic contexts involving the notions of irrealis, intention, prediction and evaluation to not facilitate the emergence of distinguishing patterns across English varieties, even though, together, these contexts do facilitate the use of mental predicates;

5. a tendency for HKE and IndE not to mirror any particular native variety in contexts involving negotiable predicated events in correspondence (in the case of HKE) and predicated events in popular writing involving opinions (in the case of IndE); and

6. a tendency for HKE and IndE not to associate these contexts of use with a specific mental predicate.

Considering the above patterns in more detail, we see with the first pattern that believe and think emerge as a lexical pair of mental predicates that associates most strongly with BrE. However, this seems to be more the case for think and less so for believe, given the closer distance between think and the variety (shown as VARIETY:GB on the plot) compared to the greater distance between the variety and believe. Generally, contexts of non-argumentativeness and positive speaker’s
evaluation seem to facilitate the uses of *think*, as illustrated in example (19), extracted from the BrE data.

(19) I think there are excellent opportunities for collaboration between us

Because of its location on the plot (i.e. further up in the quadrant), we can assume that *believe* is not only less strongly associated than *think* with the non-argumentative and positive evaluation semantic features, but also that this particular mental predicate triggers linguistic uses that contrast relatively strongly with those of *think*. However, it remains unclear what specific linguistic features contribute to the specificity of these uses given that no particular (semantic) feature is plotted in the proximity of *believe* other than its occurrence in instructional writing contexts. Ultimately, this finding is interesting for two main reasons: first, it opens the door to further exploration of the uses of *believe* and *think* in native English, particularly within the specific genres of instructional writing (where *believe* seems to yield its most characteristic uses) and editorials (where *think* seems to yield its most characteristic uses); second, even though BrE speakers appear to distinguish between the two predicates, this distinction does not seem to be salient to Asian users of English, regardless of their native linguistic background.

Moving on to the second observed pattern, *suppose* vs. *guess*, in the bottom left quadrant of Figure 1, in sharp contrast to *believe* and *think*, the semantic feature of argumentativity characterizes the usage patterns of *suppose* and *guess* – but much more so for *guess*, which is plotted much closer to the semantic feature. The argumentative use of *guess* in AmE is illustrated in (20).

(20) So I called to let you know when my bus got here – I guess you didn’t get it?

In contrast to *guess*, what characterizes the uses of *suppose* in the data is its association with speakers’ neutral evaluation of the predicated event as well as its verifiable nature, as illustrated in example (21).

(21) However, I suppose that you are much too busy to do so, having met so many brothers and sisters

Interestingly, uses such as those illustrated in (21) seem to emerge as characteristic of both the AmE and the SingE varieties, and creative writing seems to be the written genre specifically tied to those uses. However, follow-up research is necessary to establish whether, and to what extent, in this specific genre, neutral evaluation and verifiable predicated events indeed overlap in AmE and SingE. At this point, it is interesting to note the dichotomy that emerges between SingE on the one hand
and IndE and HKE on the other hand. This dichotomy is reflected in two ways: (i) in the fact that Figure 2 singles out SingE as the only ESL variety that mirrors native (AmE) usage patterns, and (ii) in the fact that, to some extent, SingE is the only ESL variety that associates strongly with one particular mental predicate, namely *suppose*. With regard to (i), although this pattern has already recently started to be documented in the ESL literature, to my knowledge, after Deshors & Gries (2016) and Horch (2016), this is only the third study so far to capture this pattern. (I will return to this finding in Section 5 and discuss its implications in the context of Schneider’s (2007) evolutionary model of language varieties.)

The fourth observed pattern involves linguistic contexts where irrealis, intention, prediction and evaluation facilitate the use of mental predicates, but are not conducive to the emergence of distinguishing usage patterns across English varieties. Although this cluster of semantic features associates neither with a particular English variety nor any specific mental predicate, its value lies in that it identifies linguistic contexts that are not particularly helpful to contrast English varieties.

With pattern 5, we are now turning our attention to the case of IndE and HKE. Starting with IndE, the top left quadrant of Figure 2 shows that in this variety mental predicates tend to occur primarily in popular, academic and student writing. In addition, mental predicates in IndE tend to (i) express an opinion (rather than a conviction, intention or prediction), (ii) contain utterances that present a proposition (rather than a speaker’s evaluation), and (iii) involve situations anchored in reality. Examples (22) to (24) illustrate uses of *think* and *believe* by Indian English users in linguistic contexts that combine opinion, proposition and reality in academic writing (22), popular writing (23) and student writing (24).

(22) I think young men are not very particular about children  
(ind:w2a-006#50:1)

(23) I believe that the 1991 census enumeration was, on the whole, a fairly good enumeration  
(ind:w2b-013#12:1)

(24) I think it is very difficult to answer this question  
(ind:w1a-001#89:2)

Continuing with IndE, because the features conviction, student writing, negative evaluation and non-negotiable predicated statement share the same quadrant as IndE (top left quadrant), we immediately associate these features with that particular ESL variety. However, given that the features are plotted close to the vertical dotted line, we can also assume that, to some extent, those features are salient in BrE and that they potentially unveil similarities between these two English varieties.
Now moving on to HKE, located at the top of the lower right quadrant of Figures 1 and 2, we observe usage trends that contrast with those observed in IndE in that in HKE, mental predicates are more frequently observed in correspondence and with negotiable predicated statements. Also, compared to IndE, a smaller number of features seem to be involved in the characterization of HKE. As a result, compared to IndE, HKE could emerge as a “blander” variety, so to speak. Overall, in light of these results, the MCA approach to mental predicates in ESL offers a promising starting point to explore the interplay of mental predicates, genres, functional uses and ESL varieties. Let us now examine the data through the lens of the classification and regression tree analysis.

### 4.2 Classification and regression tree analysis

While the MCA method unveiled a number of potentially distinguishing patterns in the uses of mental predicates by native English and ESL writers, the classification tree analysis helps us to identify which of our observed patterns are most reliable and which variables significantly affect speakers’ decision patterns in their choices of *believe*, *guess*, *suppose* and *think*. With that in mind, the classification tree analysis has confirmed that (i) there is indeed great variation in the uses of *believe*, *guess*, *suppose* and *think* across native and ESL varieties, (ii) written genres do play an important role in this variation, (iii) it is when variety and genre correlate that the lexical choices of native and ESL writers tend to differ, and (iv) not all aspects of mental predicates’ functional uses contribute equally to this variation.

Overall, the classification tree, represented in Figure 3, returned a relatively high classification accuracy of all sentences included in the data. Specifically, 79.1% were classified correctly, which is significantly different from the baseline percentage of 47.8% one would obtain by always just predicting the most frequent pattern or by just guessing randomly.

As previously mentioned (Section 3.2.2), not all variables are included in the output of a classification tree, only the ones that influence the results in a way that is statistically significant. On this basis, Figure 3 shows that out of the nine independent variables included in the study, only four turned out to be significant and to affect in some way the uses of *believe*, *guess*, *suppose* and *think*. Specifically, these variables are the following: variety, genre, argumentativity and epistemic type. This means that, at the MCA stage, even though the variables evaluation, negotiability, verifiability, epistemic class and epistemic mode helped us recognize potential linguistic contexts that are conducive to the development of characteristic uses of mental predicates in individual ESL varieties, when it comes to predicting the uses of a specific predicate across ESL varieties, these variables are
not reliable enough. In contrast, reliable predictions can be made on the basis of variety, genre, argumentativity and epistemic type. In itself, this is an interesting finding because it indicates that only two semantic features actually contribute reliably to speakers’ choices of mental predicates, namely argumentativity and epistemic type. More generally, this suggests that across English varieties, writers conceptualize mental predicates relatively similarly. What seems to be most relevant for writers, however, is more the contexts of use of these predicates rather than how their meanings are conceptualized. Returning to Figure 3, the graphical output should be read and interpreted as follows:

– one should start at the top node (node 1);
– at each node, one goes to the left or to the right to the next node (either node 2 on the left or node 11 on the right), which corresponds to assuming that the variable mentioned in the node takes on the values superimposed onto the line leading to the next node;6
– the above step is repeated until one reaches a terminal node, which provides the following information: (i) the number of the node and the number of the sentences which exhibit the combination of features by means of which one

6. In Figure 3, some of the superimposed values on the lines leading to the nodes overlap. These values are fully listed and discussed further in the text.
reached that node, and (ii) a bar plot of observed percentages of the four mental predicates (specifically, B=believe, G=guess, S=suppose, T=think).

For example, node 14 shows that there are 49 cases with a mental predicate that include an argumentative statement (path from node 1 to node 11), with a mental predicate that expresses an opinion (path from node 11 to node 12) and occurring in either academic writing, correspondence, creative writing or student writing; of these, about 85% use guess. A first glimpse at Figure 3 reveals three main general trends in the data: (i) variety only makes a difference in terms of choice of lexeme when the predicated statement is non-argumentative; (ii) when the predicated statement is argumentative, then the usage patterns of mental predicates are influenced by epistemic type; and (iii) genre only influences writers’ choice of mental predicates when these convey an opinion (as opposed to a prediction, intention or conviction). Taking a closer look at Figure 3, the data yield the significant, more fine-grained patterns described below from 1 to 8:

1. If the predicated statement is non-argumentative, if the predicate conveys the notions of intention, opinion or prediction, if the English variety is either British, Indian or Singaporean and if the mental predicate is used in academic, instructional or student writing, then the mental predicate believe is preferred 59% of the time (nodes 1, 2, 3, 4, 5), as in I believe that a married woman’s first and foremost priority is still her children and then her career (sin:w1a-010, student writing).

2. If the predicated statement is non-argumentative, if the predicate conveys the notions of intention, opinion or prediction, if the English variety is either BrE, IndE or SingE but the mental predicate is used in correspondence, creative and popular writing and editorials, then the mental predicate think is preferred 69% of the time (nodes 1, 2, 3, 4, 6), as in I think it’s because I’m more settled now and the situation in Australia isn’t that good at the moment (sin:w1b-002, correspondence).

3. If the predicated statement is still non-argumentative, if the predicate conveys the notions of intention, opinion or prediction, but if the English variety is either HKE or AmE and the genre is not creative writing, then the mental predicate think is preferred 88% of the time (nodes 1, 2, 3, 7, 8), as in I think technology develops so fast that it is beyond government’s ability to respond (usa:w2b-036, popular writing).

4. However, if the predicated statement is non-argumentative, if the predicate conveys the notions of intention, opinion or prediction and occurs in HKE or

---

7. The description of the results presented in steps 1–8 mirrors the structure used in Bernaisch et al. (2014).
AmE in creative writing, then although the preferred lexeme remains *think* (74% of the time), the other lexemes *guess* and *suppose* are also used 25% of the time (10% in the case of *guess* and 15% in the case of *suppose* (nodes 1, 2, 3, 7, 9) (this contrasts with node 8 where, in contexts that exclude creative writing, *guess* and *suppose* are hardly used but tend to be replaced by *believe*), as in *I suppose* *I’ll go back to Singapore, go back to work* (hk:w2f-012, creative writing).

5. If the predicated statement is non-argumentative and if the mental predicate conveys conviction, then all English varieties behave similarly, namely all writers have a strong preference for *believe* (75% of the time) and a less strong preference for *think* (26% of the time) (nodes 1, 2, 10), as in *I do believe that such a reductive view of language acquisition fails to fully account for the full experience of acquiring one’s first language* (usa:w1a-008, student writing).

6. If, in contrast with the above, the predicated statement is argumentative and if the mental predicate expresses the notions of conviction, intention and prediction, then genre does not influence any of the writers, who choose *guess* 83% of the time followed by *believe* (11% of the time) and *think* (5% of the time) (nodes 1, 11, 15), as in *I guess you will spend your holidays with your parents* (sin:w1b-002, correspondence).

7. However, if the mental predicate expresses an opinion, then genre makes a difference: in popular, instructional writing and editorials, all writers have a strong preference for *guess* (69% of the time), followed by *think* (23%) and *believe* (7%) (nodes 1, 11, 12, 13), as in *I would guess that the majority would be unable to answer the question* (gb:w2b-017, popular writing).

8. In contrast, in correspondence, academic, creative and student writing, *guess* is even more preferred (by all writers; 88% of the time), *suppose* is selected 8% of the time, *think* is selected 4% of the time and *believe* is not chosen at all (nodes 1, 11, 12, 14), as in *I guess she felt the need to say something* (usa:w2f-006, creative writing).

A central aspect of the above results is that variety and genre influence ESL simultaneously as well as independently depending on linguistic contexts. For both factors to affect speakers, *believe*, *guess*, *suppose* and *think* have to occur in non-argumentative contexts where an intention, opinion or prediction is conveyed. In any other contexts, the four mental predicates yield usage patterns of their own but those are not affected by the English variety of the speakers. In addition, an interesting aspect of Figure 3 is that it effectively shows how diverse the uses of the four lexemes are across linguistic settings. This diversity indicates that, despite their near-synonymy, *believe*, *guess*, *suppose* and *think* should not be assimilated too quickly. For instance, based on Figure 3, significant variation in the use of *believe*
and think is only observed in contexts that are non-argumentative; and within such contexts, both lexemes are anchored in different linguistic settings: that is, believe is the preferred item when conviction is expressed and when opinion, intention or prediction is expressed but then only in BrE, IndE and SingE within the academic, instructional and student writing genres. In all other non-argumentative contexts, think is the lexeme of choice. This pattern contrasts with argumentative contexts where believe and think are the least preferred lexemes and where guess is clearly the preferred choice. Ultimately, while the higher frequency of guess in argumentative contexts was already observed at the MCA stage, the classification tree not only confirms this finding but also specifies it by adding that the finding is true regardless of genre and epistemic type. Finally, while, across the board (i.e. in argumentative and non-argumentative contexts), suppose is the lexeme the least opted for, it is interesting to note that it tends to be most frequently used in creative writing. This suggests that this particular genre is an ideal candidate to capture the usage patterns that are characteristic of suppose.

5. Discussion and concluding remarks

This study set out to investigate the uses of the mental predicates believe, guess, suppose and think across native and non-native written corpora of English. The focus was on exploring to what extent native and Asian speakers of English structure the four mental predicates in question differently and to what extent different genres and semantic functions affect their uses. Methodologically, I adopted a qualitative-quantitative corpus-based approach including a qualitative analysis of the data (based on an annotation scheme developed in Krawczak 2014) and two statistical techniques, namely an exploratory MCA and a confirmatory CART analyses. Altogether, this methodological approach helped me to address the following research questions: (i) to what extent do the patterns of use of believe, guess, suppose and think differ across native and ESL varieties? (ii) to what extent can cross-varietal variation in the use of these mental predicates be observed across different written genres? and (iii) to what extent can native and ESL writers be observed to structure the meaning of believe, guess, suppose and think differently? Overall, with this methodological approach, I was able to draw a comprehensive picture of the variation patterns of epistemic stance in Asian Englishes.

With regard to (i) and (iii), the present study showed that the semantic structure of believe, guess, suppose and think is an aspect of language (use) that is a fertile breeding ground for the development of variation patterns across native English and ESL and across Asian Englishes. Indeed, unpacking the semantic load of mental predicates (in terms of epistemic type, class and mode and in terms of
argumentativity, verifiability, evaluation and negotiability) made it possible to pinpoint the specific semantic facets of mental predicates that are salient to native and Asian English writers. The process of semantic decomposition therefore provided a useful new perspective on cross-variety variation (‘new’ in the sense of approaching this variation from the perspective of semantic structure, as opposed to syntactic variation which has recently become one of the most studied phenomena in ESL corpus-based research). Focusing on semantic structure revealed that the uses of mental predicates by different populations of English writers vary more when semantic features such as intention, opinion and prediction are involved compared to a feature like conviction. In a similar vein, although argumentative contexts lead to different usage patterns of the four predicates in focus, those contexts are not conducive to the development of distinctive patterns across English varieties and across Asian Englishes. Ultimately, this finding shows that the patterns of use of believe, guess, suppose and think differ across native English and ESL varieties at a relatively abstract semantic level and that in order to capture variation patterns, it is necessary to adopt methodological techniques that allow analysts to break down the structure of lexical items and explore their variation across dialects. In turn, such methods made it possible to reach an unprecedented level of granularity when exploring semantic variation across ESL varieties. This is essentially based on the finding that mental predicates conveying the notions of intention and prediction yield similar usage patterns regardless of whether or not the predicated statement is argumentative or not (unlike mental predicates that express conviction and opinion and whose uses do vary based on argumentativity).

In addition, the current analysis helped us to identify an important distinction between the native and the Asian English varieties in relation to the semantic feature of negotiability. As observed in Krawczak (2014), this feature plays an important role in distinguishing between think and believe on the one hand and suppose and guess on the other hand. Indeed, she notes, in native English “[a] property that points towards the semantic completeness of think and believe is their dissociation from non-negotiability. […] The picture changes considerably when we turn to guess and suppose […] the two expressions exhibit usage properties of semantic bleaching, which is characteristic of subjective construal” (Krawczak 2014: 322). She adds that across BrE and AmE, “although it is possible to identify some subtle differences across the dialectal divide, in fact, the distinctions revealed are not very conspicuous”. Interestingly, when Asian Englishes are taken into account, the semantic feature of negotiability becomes irrelevant and does not help us capture usage patterns characteristic of Asian English writers. However, a feature that does make a difference in separating believe and think from suppose and guess, and that Krawczak (2014) does not take into account, is genre. In fact, in the present work, genre emerged as the main feature with which we are able to contrast guess and
suppose (see nodes 13, 14 and 15 in Figure 3). So while Krawczak’s (2014) study uses blog-based data, a richer type of data dialogically speaking, the current analysis suggests that writing (in ICE) is quite variable in itself. Further, the current study is a case in point showing how important it is to design analyses that account for register (variation).

With regard to (ii), to what extent cross-varietal variation is observed across individual written genres, as already mentioned above, both methodological techniques (MCA and CART) unveiled a significant correlation between English variety and genre. Specifically, in non-argumentative contexts where the notions of intention, opinion and prediction are expressed, IndE and SingE were found to yield patterns similar to BrE; however, in the exact same semantic contexts, HKE was found to resemble AmE more. Because of the high degree of granularity that the CART technique allowed us to reach, it was possible to take a close look at the preference patterns of individual mental predicates in specific genres. For instance, correspondence, creative writing, popular writing and editorials emerged as specific genres where think is preferred by BrE, IndE and SingE writers. In contrast, in the same three varieties, believe is the most frequently used predicate in academic and instructional writing. When HKE writers use mental predicates in correspondence, however, there is a much higher chance that they will resort to think rather than believe; and the chances that they resort to guess and suppose are negligible. Based on these results, it is important that follow-up studies continue to account for the variety-genre correlation so as to draw a fine-grained picture of mental predicates in Asian Englishes, a picture that is as realistic as it can be.

Beyond the above-discussed research questions, the present study brings to light an interesting aspect of ESL varieties that is only just starting to emerge in the relevant literature, that is, the question of directionality in the process of emancipation of ESL varieties in the late 20th and early 21st century. In this regard, Asian Englishes provide an interesting point of discussion leading to a possible reassessment of what it means for a variety with a British colonial historical background to advance in its emancipation process. For the past 30 years or so, much of the body of research on ESL varieties has focused on trying to emancipate ESL varieties from their historical input varieties. Theoretically, this body of research is anchored in Schneider’s (2007) dynamic-evolutionary model. As discussed at the onset of the present paper (Section 2.3), the model assumes that ESL varieties follow a five-stage development process towards nativization and that HKE (transitioning from phase II to III) is less advanced than IndE (transitioning from phase III to IV), which is less advanced than SingE (at an advanced stage of phase IV). Based on this model, one would intuitively anticipate that the more advanced variety (i.e. SingE) would be the least similar to a native variety (particularly BrE as the historical source variety), compared to the less advanced
IndE and HKE varieties. Unexpectedly, however, recent studies such as Deshors & Gries (2016) have started to question this emancipation pattern. Focusing on verb complementation constructions, Deshors & Gries (2016) find that, although SingE was expected to be the least similar to native English, the variety resembles AmE more, and IndE/HKE do not resemble any particular native variety. This is a very interesting finding because, while work such as Leimgruber (2011) shows that SingE is undergoing a language shift from (predominantly) second-language to (increasingly) native language, one could envisage that SingE’s development is more complex than it may appear. Indeed, its emancipation process from BrE to (near-)native variety could possibly only be one facet of its development. A second facet could involve an assimilation process of AmE patterns due to linguistic influential forces on a world scale. In other words, the above results raise the question of a possible Americanization of SingE.

Focusing on the CART output, the notion of assimilation of AmE patterns in the context of worldwide linguistic influences can be taken one step further with HKE. As observed in Figure 3, HKE, the least advanced variety, is most similar to AmE. In turn, this raises the question whether, depending on their evolutionary stages, Asian Englishes could follow different paths of development. According to this line of argumentation, varieties at the least advanced stage of development could be more receptive to contemporary influential cultures such as the American culture, as there is no apparent reason to believe that HKE, IndE and SingE users would be exposed to more or less AmE. Yet, only HKE users are observed to yield usage patterns that are similar to AmE. While, admittedly, the MCA and CART results in the current study are not totally in line with one another, they nonetheless raise the important question of how, by being exposed to (arguably) more dominant native English varieties such as AmE, speakers of Asian English collectively shift the development of their English variety and how this shift should be accounted for in the broader theoretical context of explaining the development of English dialects around the world. In order to elucidate these questions, more systematic explorations that contrast ESL varieties, BrE and AmE are needed.

In sum, as one of the first corpus-based analyses to contrast ESL varieties semantically using multivariate methodology, the present study has shown the benefits of complementing existing ESL studies anchored in lexico-grammar with investigations focusing on patterns of lexicalization. The results have shown that in order to make sense of semantic patterns it is important to account for extra-linguistic factors such as genre, as different ESL writers tend to structure the meaning of believe, guess, suppose and think differently depending on their type of writing. Ultimately, although the present study should be replicated on the basis of larger data sets including a wider range of ESL varieties, the results raise
important questions with regard to the emancipation process of Asian Englishes and the place that semantic structure holds in this process.

Acknowledgments

I wish to thank Gaëtanelle Gilquin, Sandra Götz, Tobias Bernaisch and two anonymous reviewers for their thorough and valuable feedback on an earlier version of this paper.

References


Author’s address

Sandra C. Deshors
Department of Languages & Linguistics
New Mexico State University
MSC 3L, Las Cruces, NM 88003
USA
sandracdeshors@gmail.com