

Article

Reassessing the Learner Englishes–New Englishes Continuum: A Lexico-Grammatical Analysis of TAKE in Written and Spoken Englishes

Yating Tao  and Gaëtanelle Gilquin * 

Centre for English Corpus Linguistics, Université catholique de Louvain (UCLouvain), 1348 Louvain-la-Neuve, Belgium; yating.tao@uclouvain.be

* Correspondence: gaetanelle.gilquin@uclouvain.be

Abstract

This study reexamines Learner Englishes (LEs)–New Englishes (NEs) continuum by considering intervarietal variation, mode differences, and multiple linguistic levels. Relying on comparable written and spoken corpus data, we investigate the valency patterns and senses of the verb TAKE across two LEs (Mainland Chinese English (MCE) and Belgian French-speaking English (BFE)) and two NEs (Singapore English (SgE) and Hong Kong English (HKE)) within the Extra- and Intra-territorial Forces (EIF) Model. The study examines whether internal linguistic factors, namely, mode (writing and speech) and linguistic levels (valency patterns and senses), influence the variety positioning along the LEs–NEs continuum and whether this positioning reflects the expected proximity cline to native English (NativeE) (BFE > MCE > HKE > SgE) established within the EIF Model. Our quantitative results reveal that individual varieties intermingle depending on mode and linguistic levels rather than occupying stable positions along the LEs–NEs continuum. Dendrogram analyses yield distinct variety clustering patterns that contradict the expected proximity cline to NativeE. Qualitatively, we identify some shared linguistic features across LEs and NEs that suggest common underlying language learning strategies. These results contribute to variationist linguistics by demonstrating that English varieties exhibit dynamic development trajectories shaped by language-internal factors (e.g., mode and linguistic levels). We propose refining the EIF Model to incorporate language-internal dimensions, thereby bridging the gap between LEs and NEs through a more nuanced theoretical framework.

Keywords: Learner Englishes–New Englishes continuum; extra- and intra-territorial forces model; mode; valency patterns; senses



Academic Editor: Julia Davydova

Received: 18 July 2025

Revised: 23 October 2025

Accepted: 4 November 2025

Published: 13 November 2025

Citation: Tao, Y., & Gilquin, G. (2025).

Reassessing the Learner Englishes–New Englishes Continuum: A Lexico-Grammatical Analysis of TAKE in Written and Spoken Englishes.

Languages, 10(11), 285. <https://doi.org/10.3390/languages10110285>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

([https://creativecommons.org/](https://creativecommons.org/licenses/by/4.0/)

<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Learner Englishes (LEs) and New Englishes¹ (NEs) were traditionally treated as two distinct categories, which can be reflected in Kachru's (1985) Three Concentric Circles, where LEs and NEs correspond to the expanding circle and the outer circle, respectively. LEs in the expanding circle developed through the global extension of English as an international language of communication (i.e., globalization), as in China and Belgium, where English has no official status and mainly serves for international communication beyond classroom settings.² By contrast, NEs in the outer circle were mainly developed due to British or American colonial history, as in Singapore and India, where English is an official or semi-official language and serves for both intranational and international communica-

tion across a wide range of domains (e.g., education, administration, and literature). In addition, LEs and NEs were labelled, respectively, as norm-dependent and norm-developing in Kachru's (1985, p. 17) model due to their different norm-orientations: exonormative for LEs vs. endonormative for NEs. To be more specific, LEs speakers usually orient towards native English (NativeE) (mainly British English (BrE) or American English (AmE)), while NEs speakers are thought to have the potential to develop local standards and norms (Mukherjee & Hundt, 2011, p. 2). Given these differences, LEs and NEs have long been addressed in two separate linguistic fields: second language acquisition (SLA) research for LEs and contact linguistics for NEs.

This rigid distinction between LEs and NEs is problematic, however, considering that there are obvious parallels between the two. First, both types of varieties share the non-native English status and involve language contact in a multilingual context. Second, they are both increasingly subject to global influences through digital mass media and new communication technologies, which provide exposure to diverse global discourse communities that transcend traditional geographical and historical distinctions between LEs and NEs, though with different degrees. Third, LEs and NEs have been claimed to be affected by similar cognitive mechanisms of language acquisition (Lowenberg, 1986; Williams, 1987; Biewer, 2011; Schneider, 2012), such as overgeneralization, redundancy, and analogy. These shared cognitive mechanisms inevitably lead to some common non-standard linguistic features in LEs and NEs. For instance, studies on phrasal verbs with *up* reveal that both LEs (Gilquin, 2015b) and NEs (Zipp & Bernaisch, 2012) tend to insert *up* after the verb as a superfluous particle, as in *rise up* and *cope up with*. These non-standard features are, however, usually referred to as "errors" by LEs researchers and "innovations" by NEs researchers (see Gilquin, 2015a; Deshors et al., 2018 for further discussion). This implies that, in such cases, the terminological difference is crucially established based on researchers' disciplinary perspective rather than on inherent linguistic levels. To overcome this limitation, a more neutral term, "non-standard features", is used here to describe usages in native and non-native English varieties that diverge from established expert British or American norms.

The disconnect between the two fields, therefore, cannot reflect the linguistic reality. Sridhar and Sridhar (1986) criticized this separation as a "paradigm gap". They advocated for an integrated approach to the two types of varieties to promote the rapprochement between the two paradigms. Despite this early call for an integration, it was not until 2008 that a growing body of research began to address the paradigm gap, mainly by comparing LEs and NEs using corpus resources (e.g., Nesselhauf, 2009; Gilquin, 2011; Mukherjee & Hundt, 2011; Davydova, 2012). Notably, several workshops³ were organized to encourage scholars from both fields to reevaluate the links between LEs and NEs. These research endeavours generated significant attention and stimulated further dialogue among linguists in the following years. This is evident in recent publications, including several monographs (e.g., Buschfeld, 2013; Edwards, 2014; Percillier, 2016; Davydova, 2019) and articles in relevant journals, which have contributed to bringing the two paradigms closer together. One important finding from these corpus studies is that the distinction between LEs and NEs is not clear-cut but represents a continuum with many in-between categories (e.g., Gilquin & Granger, 2011; Buschfeld, 2013; Edwards, 2014). This advancement has inspired researchers to reflect on the validity of existing NEs models in LEs contexts, and vice versa, and to develop joint theoretical frameworks that integrate LEs and NEs, such as Buschfeld and Kautzsch's (2017) Extra- and Intra-territorial Forces (EIF) Model.

While the EIF Model provides a valuable framework to capture the dynamic development trajectories of individual varieties along the LEs-NEs continuum, its focus on external sociolinguistic forces leaves open important questions about the role of language-internal

factors. In view of this, this study aims to extend the EIF Model by considering intervarietal variation, mode differences (writing and speech), and multiple linguistic levels (valency patterns and senses), which are largely neglected in the existing comparative studies of LEs and NEs (see Section 3). Two LEs (Mainland Chinese English (MCE) and Belgian French-speaking English (BFE)) and two NEs (Singapore English (SgE) and Hong Kong English (HKE)) are chosen for this purpose, as they are expected to occupy different positions along the continuum within the EIF Model (see Section 2).

This article is organized as follows. Section 2 discusses the rationale of the LEs-NEs continuum and introduces the EIF Model, establishing the hypothesized proximity of LEs and NEs to NativeE within the EIF Model. Section 3 provides a systematic summary of existing corpus-based comparative studies of LEs and NEs and the limitations of these studies. Our research questions are outlined at the end of this section. Section 4 presents the data and methods used in this study. Our quantitative and qualitative results are provided and discussed in Section 5. Finally, Section 6 closes the article by providing some conclusions.

2. LEs-NEs Continuum and the EIF Model

As pointed out above, the traditional categorical distinction between LEs and NEs has gradually been replaced by a continuum that recognizes many hybrid varieties with characteristics of both types of varieties in between. This shift is grounded in empirical evidence from researchers in contact linguistics and SLA research. For instance, some researchers observed that NEs are not necessarily the result of British or American colonial history (e.g., Buschfeld & Kautzsch, 2014 on Namibian English; Mežek, 2024 on Swedish English, to name just a few), but arise primarily from the general force of globalization, which leads to the increasing use of English across multiple social and institutional domains in those regions and thus makes them more similar to NEs than to LEs (e.g., English is used for intranational communication). Furthermore, English in previous colonies is dynamic and may change from one status to another flexibly. For example, some scholars argue that shifts have occurred from NEs to LEs status in the case of Cyprus (Buschfeld, 2013) and Hong Kong (Görlach, 2002, pp. 109–110). These results point to the invalidity of colonial history as the sole criterion for determining the variety status.

In addition, researchers noticed a continuum between LEs and NEs as they found that the two types of varieties exhibit many shared non-standard linguistic features, presumably reflecting similar underlying cognitive mechanisms of language acquisition (e.g., Biewer, 2011; Callies, 2016; Nesselhauf, 2009) and recent sociolinguistic practices such as the use of digital media (e.g., Buschfeld & Kautzsch, 2017). For instance, Nesselhauf (2009), as one of the earliest researchers comparing LEs and NEs, examined a range of lexicogrammatical co-selection phenomena in four LEs (German, French, Finnish, and Polish) and four NEs (India, Singapore, Jamaica, and Kenya) and identified a number of new prepositional verbs that are shared across LEs and NEs, including *discuss about*, *demand for*, and *enter into*. These new prepositional verbs suggest that similar cognitive processes are at play in both LEs and NEs, i.e., “nativized semantico-structural analogy” (Mukherjee & Hoffmann, 2006). The prepositional verb *discuss about*, for example, can be an analogy to *a discussion about* or *speak/talk about*.

More recent multifactorial analyses (e.g., Deshors, 2014; Gilquin & Meriläinen, 2024) provide additional evidence for the LEs-NEs continuum. For instance, in her study of the dative alternation of the verb GIVE (prepositional vs. ditransitive construction) across LEs and NEs, Deshors (2014) reported that German LE clusters with NEs, rather than with French LE, in both constructions. This indicates that non-native English varieties do not necessarily cluster together according to their LEs or NEs status but intermingle along the continuum. This intermingling clustering challenges the prevailing LEs-NEs continuum

model, which posits a linear and gradual progression from LEs to NEs, echoing [Gilquin and Granger's \(2011, 2021\)](#) finding of a non-linear variety relationship.

The increasing blurring between LEs and NEs and the growing force of globalization on the development of both types of varieties have motivated researchers to question whether existing NEs models could account for LEs contexts and seek to adapt or develop models to capture these shifts. For instance, [Schneider \(2014\)](#) tested the applicability of his Dynamic Model of postcolonial Englishes (PCEs) ([Schneider, 2007](#)) to LEs territories, including China, (South) Korea, and Rwanda, and found that “the Dynamic Model is not really, or only to a rather limited extent, a suitable framework to describe this new kind of dynamism of global Englishes” ([Schneider, 2014](#), p. 28). Against this background, he proposed the concept of “Transnational Attraction”, which refers to “the appropriation of (components of) English(es) for whatever communicative purposes at hand, unbounded by distinctions of norms, nations or varieties” ([Schneider, 2014](#), p. 28). In the same spirit, [Edwards \(2014\)](#) applied the Dynamic Model to the Netherlands to explore the status of English there. She found that it is difficult to classify English in the Netherlands as either LE or NE because it displays characteristics of both types of varieties, and she called for an integrated model for LEs and NEs.

In response to [Schneider \(2014\)](#) and [Edwards \(2014\)](#), [Buschfeld and Kautzsch \(2017\)](#) designed the EIF Model, which demonstrates the parallels but also explains the differences between LEs (i.e., non-postcolonial Englishes (non-PCEs)) and NEs (i.e., PCEs) within a unified framework (see Figure 1). The EIF Model builds on two essential foundations. First, it operationalizes [Schneider's \(2014\)](#) abstract notion of “Transnational Attraction” through a defined set of extra- and intra-territorial forces that drive the development of both LEs and NEs. The external forces include (a1) colonization, (a2) language policies, (a3) globalization, (a4) (extra) foreign policies, and (a5) sociodemographic background of a country. The internal forces comprise (b1) attitudes towards colonizing power, (b2) language policies/language attitudes, (b3) “acceptance” of globalization, (b4) (intra) foreign policies, and (b5) (intra) sociodemographic background of a country ([Buschfeld & Kautzsch, 2017](#)). The external forces (a) and internal forces (b) correspond to each other and work as two parallel lines throughout all phases ([Buschfeld & Kautzsch, 2017](#), p. 114). The first force, colonization, is the only factor that could differentiate NEs from LEs ([Buschfeld & Kautzsch, 2020](#)), as there is no colonization history in LEs contexts. The second foundation of the EIF Model is that it adapts [Schneider's \(2007\)](#) diachronic development of varieties along five phases for both LEs and NEs contexts: (1) foundation, (2) exonormative stabilization, (3) nativization, (4) endonormative stabilization, and (5) differentiation.

A particularly relevant aspect of the EIF Model to the present study is its conceptualization of the LEs-NEs continuum, which serves as the backbone of the EIF Model. This continuum represents a developmental trajectory from LEs to NEs and potentially native status ([Buschfeld & Kautzsch, 2017](#), p. 117), which corresponds to the traditional categorization of English as a foreign language (EFL), English as a second language (ESL), and English as a native language (ENL), as illustrated in Figure 1. However, the model emphasizes that such developments are not unidirectional only (see bidirectional arrows in Figure 1), allowing individual varieties to move in either direction along the continuum. Despite its directional flexibility, the development trajectory in the EIF Model implies a predictable proximity cline to native norms. As observed by some researchers (e.g., [Schneider, 2004](#); [Mukherjee & Gries, 2009](#)), varieties at later developmental phases (endonormative stabilization and differentiation) should distance themselves from NativeE, as more advanced varieties in [Schneider's \(2007\)](#) Dynamic Model are more divergent from native norms (e.g., BrE or AmE). Given that the EIF Model largely builds on the Dynamic Model, we can expect that developmental phases display an inverse relationship with the proxim-

ity to native norms within the EIF Model. In light of this, LEs should be closer to NativeE than NEs are.

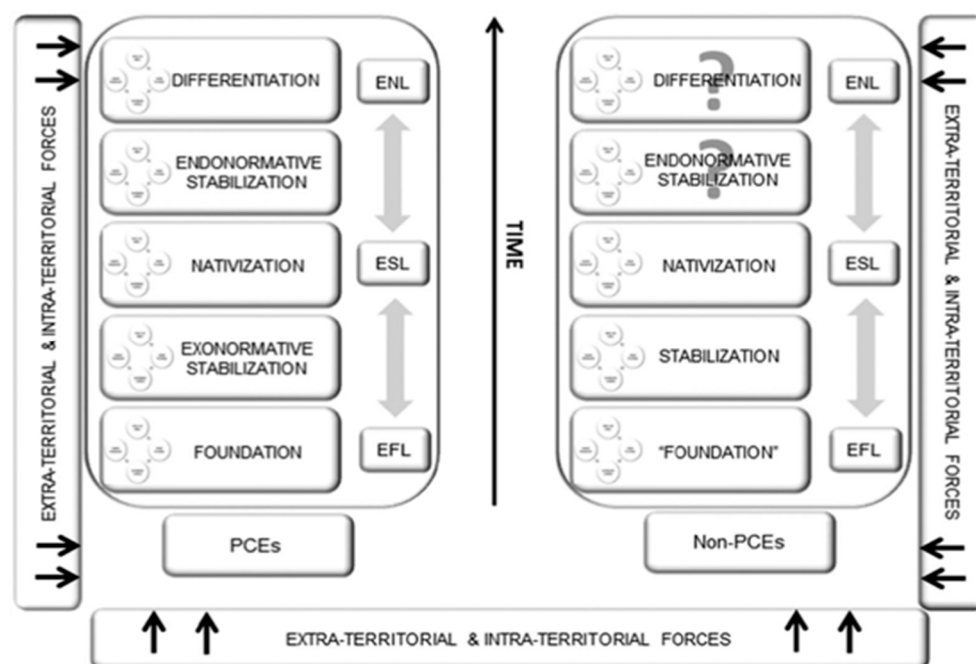


Figure 1. The EIF Model (reproduced from Buschfeld & Kautzsch, 2017, p. 117).

Our four varieties at different developmental phases may suggest a potential proximity cline to NativeE. SgE is predominantly in phase 4 (endonormative stabilization) but shows early signs of phase 5 (differentiation), whereas HKE is primarily in phase 3 (nativization) but retains some traces of phase 2 (exonormative stabilization) (Schneider, 2007). MCE, despite being conventionally classified as an LE, exhibits traces of nativization (phase 3), including distinctive features (e.g., Bolton, 2003; Xu, 2010; Ma & Xu, 2017), early codification efforts (Xu, 2010), and acceptance in Chinese bilingual literary works (e.g., *The Importance of Living* by Lin Yutang and *Red Dust* by Ma Jian). This has led some researchers to argue that MCE is in the transforming stage of nativization (phase 3) (Ma & Xu, 2017, p. 195). Nevertheless, researchers usually place MCE closer to the LEs end on the LEs-NEs continuum (e.g., He & Li, 2009; Gilquin, 2024). It can therefore be argued that MCE is primarily an LE but displays early traces of phase 3 (nativization). BFE, by contrast, maintains clear LE characteristics, with English use being mainly restricted to education and international contexts (Meunier, 2020) (i.e., positioned at the earliest developmental phase). Following the EIF Model, developmental phases (SgE > HKE > MCE > BFE) should display an inverse relationship with the proximity to native English norms, yielding the following predictable proximity cline: BFE > MCE > HKE > SgE. However, recent research challenges this linear progression and suggests that individual varieties intermingle along the LEs-NEs continuum rather than clustering according to their assigned variety status (e.g., Deshors, 2014; Edwards & Lange, 2016). We therefore examine whether LEs and NEs cluster based on their assigned variety status ($\{BFE > MCE\} > \{HKE > SgE\}$) or intermingle along the continuum.

It should be noted, however, that other clines could be predicted based on the literature. The exposure-based hypothesis (e.g., Gilquin, 2016a, 2025) within the usage-based theory of language acquisition (e.g., Eskildsen, 2009; Diessel, 2014) might thus lead one to expect that NEs should be more similar to NativeE than LEs are, as exposure to English is generally more extensive in NEs contexts. This theoretical tension, therefore, necessitates

empirical studies to investigate whether the EIF-based proximity cline can account for the observed patterns.

Furthermore, recent refinements to the EIF Model introduce additional complexity to the proximity cline. Buschfeld et al. (2018, p. 24) improved the EIF Model by introducing a third dimension besides extra- and intra-territorial forces and developmental phases, namely, “variety-internal heterogeneity”, which emphasizes the role of sociolinguistic variables such as age, ethnicity, and social status on the position of individual LEs and NEs along the macro-level continuum. This theoretical advancement recognizes the dynamic development trajectories of LEs and NEs along the continuum, which is constrained by a set of sociolinguistic factors. Empirical studies have corroborated this dynamic trajectory. For instance, Buschfeld (2013) showed that the older generation in Cyprus exhibits more NEs-like characteristics through a higher frequency of nativized linguistic features (e.g., double past tense marking), while the younger generation shows more LEs-like characteristics. This suggests that different age groups within the same variety can occupy different positions along the LEs-NEs continuum. Similarly, Davydova (2012) demonstrated how socioeconomic settings systematically influence the linguistic features in Russia and India, with speakers from smaller cities producing more basilectal forms, while speakers from large cities produce more acrolectal forms. This illustrates how socioeconomic factors lead to systematic intra-varietal differences that may interact with the positioning of individual varieties along the continuum.

While the above studies have provided empirical evidence for the role of external sociolinguistic factors in the positioning of a variety along the LEs-NEs continuum, it remains unclear whether language-internal factors might influence the position of individual varieties along the continuum. A small body of research has noticed that different linguistic levels may influence the positioning of varieties along the LEs-NEs continuum (e.g., Edwards & Laporte, 2015). However, this crucial perspective has not received much attention and lacks comparative studies of LEs and NEs examining additional language-internal factors.

3. Comparing NEs and LEs: Existing Corpus Research

Corpora representing LEs and NEs, especially the International Corpus of Learner English (ICLE) (Granger et al., 2020) and the International Corpus of English (ICE) (Greenbaum, 1988), have facilitated systematic and rigorous comparisons between the two types of varieties as Sridhar and Sridhar (1986, p. 12) advocated. Over the years, a range of linguistic phenomena have been investigated in this context, including phonetic features (e.g., Götz, 2015), lexical uses (e.g., Callies, 2016; Gilquin, 2024), phraseology (e.g., Nesselhauf, 2009; Gilquin, 2011; Edwards & Laporte, 2015), pragmatic markers (e.g., Gilquin, 2016b; Davydova, 2019), and stylistics (e.g., Bernaisch & Götz, 2021). Furthermore, a few studies have analysed several linguistic phenomena to offer a more holistic view (e.g., Szmrecsanyi & Kortmann, 2011; Gilquin, 2015a; Percillier, 2016). All these studies highlight the feasibility and value of comparing two types of varieties by illuminating their (dis)similarities and identifying the factors that influence their degree of similarity, such as exposure to English (Gilquin, 2016a) and English proficiency levels (Laporte, 2012). The accumulating evidence from these studies has led most researchers to support the gradual LEs-NEs continuum discussed above. However, a small number of researchers hold the exact opposite opinion, advocating for the traditional LEs/NEs dichotomy (e.g., Götz & Schilk, 2011; Szmrecsanyi & Kortmann, 2011; Paulasto & Meriläinen, 2023). Gilquin and Meriläinen (2024) attribute this ongoing debate to the different linguistic phenomena investigated and the methods chosen in previous studies. In addition, this debate might be related to some limitations in the existing studies that might undermine the efforts to establish a reliable

relationship between LEs and NEs, such as mode differences, genre differences, and aggregate approaches.

The first limitation arises from **mode differences** (writing and speech). On the one hand, many studies only focus on one mode, especially writing. This skewness is largely due to the overall challenges of collecting and transcribing spoken data. Even rarer are studies that combine spoken and written data to investigate linguistic features across LEs and NEs. However, a systematic LEs-NEs comparison requires both types of data, as mode is an important factor that influences language use (Biber, 1991, p. 47). Notable exceptions are Gries and Deshors (2015) and Meriläinen (2017), who investigate the dative alternation and progressive forms, respectively, across LEs and NEs using both written and spoken data. On the other hand, some studies directly compare written data from one group with spoken data from another (e.g., Hilbert, 2011; Paulasto & Meriläinen, 2023). For example, when investigating the interrogative inversion, Hilbert (2011) used spoken data from ICE representing SgE and Indian English while relying on written data from the Hamburg Corpus of Irish English to represent Irish English.

The second limitation to corpus comparability is related to **genre differences**, which often arise from the difficulty of finding comparable LEs and NEs corpora. For instance, to investigate (dis)fluency features in LEs and NEs, Götz (2015) compared the private conversations and broadcast discussions of the Sri Lankan component of ICE with interviews of German English learners from the Louvain International Database of Spoken English Interlanguage (LINDSEI) (Gilquin et al., 2010). While this choice can be partly explained by the scarcity of comparable datasets, it may lead to genre-related effects that complicate the interpretation of the results.

Another possible limitation is the **aggregate approach**, which treats LEs or NEs as one group without accounting for possible intervarectal and intravarietal variation, such as first language (L1) influence (e.g., Gilquin, 2015a; Nesselhauf, 2009). This approach may raise questions about whether the observed differences or similarities are merely explained by assigned variety status. For instance, in her investigation of several co-selection phenomena, Nesselhauf (2009) aggregated four components of ICLE (L1 German, French, Finnish, and Polish) as ICLE-4L1 to compare them with several NEs from ICE, without further discussing the possible L1 influence. Such L1 influence has, however, been highlighted by Laporte (2012), who reported different pictures for the distributions of non-standard features of MAKE-patterns across groups (LEs vs. NEs) and varieties. This suggests that different populations with different L1s may have different tendencies to use non-standard features, thereby highlighting the potential problem of the aggregate approach and the importance of also carefully examining each variety individually.

Bearing these limitations in mind, this study aims to contribute to the collective corpus efforts of bridging the paradigm gap between SLA research and contact linguistics by comparing LEs and NEs. Special focus will be placed on the lexis-grammar interface, which has been shown to be prone to the emergence of new linguistic features (Schneider, 2007). For this purpose, we investigate the valency patterns and senses of TAKE across two LEs (MCE and BFE) and two NEs (SgE and HKE), using NativeE as a reference, based on comparable written and spoken corpus data. We are especially interested in whether language-internal factors, particularly mode differences (writing and speech) and multiple linguistic levels (valency patterns and senses of TAKE), influence the positioning of individual varieties along the LEs-NEs continuum. By “linguistic levels”, we mean the different dimensions according to which language use can be examined and, in this study, we specifically refer to the structural level (valency patterns of TAKE) and the semantic level (senses of TAKE). The research questions can be specified as follows:

1. Does the positioning of the individual LEs and NEs along the continuum differ between written and spoken modes? Does this support the expected proximity cline to NativeE (BFE > MCE > HKE > SgE)?
2. Does the positioning of the individual LEs and NEs along the continuum vary according to the valency patterns and senses of TAKE? Does this support the expected proximity cline to NativeE (BFE > MCE > HKE > SgE)?
3. Are there shared non-standard features of valency patterns and senses of TAKE across LEs and NEs? What are the possible motivations for them?

4. Data and Methods

4.1. Corpus Data

To control for the possible mode and genre effects discussed above, this study investigates both written data (limited to student argumentative essays) and spoken data (limited to informal interviews) to ensure corpus comparability across LEs and NEs. The details of the selection process are presented below.

The written data for the target LEs and NEs were drawn from three main resources: ICLE, ICE, and the International Corpus Network of Asian Learners of English (ICNALE) (Ishikawa, 2023). ICLE consists of (mainly) argumentative essays in LEs produced by university students from 25 mother tongue backgrounds, with a higher intermediate to advanced proficiency level in English. ICE provides a wide range of written and spoken genres in native varieties (e.g., BrE) and NEs. ICNALE includes topic-controlled spoken and written data produced by novice to advanced university students in Asian countries or regions.

The MCE written data were student essays drawn from the Chinese components of ICLE (ICLE-MC) and ICNALE (ICNALE-MC). The Chinese component of ICLE was collected from two groups (Granger et al., 2020, p. 37): Cantonese writers from Hong Kong (around 90%) and Chinese writers (mainly Mandarin writers) from mainland China (about 10%). Two main criteria were adopted to select ICLE-MC: students' native language (Chinese Mandarin or Chinese) and the institution where the data were collected (University of Portsmouth, UK). Due to the limited size of ICLE-MC, it was supplemented with student essays from ICNALE-MC (ICNALE-MC-stw). To ensure comparability with ICLE in terms of proficiency levels, only essays written by upper-intermediate and advanced students in ICNALE-MC (i.e., only B1+ and B2+) were selected (Ishikawa, 2023, p. 27).

For the BFE written data, essays produced by French-speaking Belgian students in ICLE were used (ICLE-FR).

The SgE written data were student essays drawn from both ICE and ICNALE. As the limited size of student essays and examination transcripts from ICE (ICE-SIN-stw), approximately 50,000 words for each variety, cannot provide adequate data for detailed analyses (only 57 cases of TAKE), additional data were necessary. We therefore included student essays from the Singapore component of ICNALE (ICNALE-SIN-stw). Different from ICNALE-MC, ICNALE-SIN starts from the B1+ level, so all essays could be included without jeopardizing comparability with ICE-SIN-stw.

For the HKE written data, essays were sourced from the Hong Kong component of ICLE (ICLE-HK). Only essays written by students with Cantonese as a native language from the Hong Kong Polytechnic University were included. These texts provide sufficient data for HKE and are highly comparable to the data of other selected varieties.

The Louvain Corpus of Native English Essays (LOCNESS) (Granger, 1998) served as a native baseline to provide native written control data for comparing LEs and NEs. Only the American section of it was used because the British section contains too limited data (less than 100,000 words) compared to LEs and NEs written data. Additionally, combining the

British and American components of LOCNESS might introduce undesirable variability in the reference corpus, given that these two varieties differ from each other in terms of certain linguistic features (see [Algeo, 2006](#); [Baker, 2017](#)).

While several corpora were combined to obtain adequate written data, the spoken data come from three highly comparable corpora in this study: LINDSEI ([Gilquin et al., 2010](#)), the New Englishes Student Interviews (NESSI) corpus ([Gilquin, 2024](#)), and the Louvain Corpus of Native English Conversation (LOCNEC) ([De Cock, 2004](#)).⁴ These three “sister corpora” are replicas of each other, designed according to the same criteria, and therefore serve as important resources for comparing LEs and NEs. Each corpus comprises around 50 informal interviews with university students in the respective countries or regions, following the same structure with three different tasks: a short monologue on one of three set topics, a free informal discussion, and a description of the same sequence of pictures. All interviews were transcribed following the same transcription guidelines.

For this study, the Chinese (LINDSEI-MC) and French components (LINDSEI-FR) of LINDSEI were used for LEs spoken data, while the Singapore (NESSI-SIN) and Hong Kong components (NESSI-HK) of NESSI were used for NEs spoken data. LOCNEC, which contains 50 informal interviews with British university students, provided native spoken control data to compare LEs and NEs speech. It should be noted that only spoken data produced by interviewees from the three corpora are analysed, as the focus of this study is specifically on students’ speech.

We acknowledge that there are differences between ICNALE and ICLE or ICE, which might influence the distribution of valency patterns and senses of TAKE in writing across varieties. ICNALE was compiled based on two topics: “It is important for college students to have a part-time job” and “Smoking should be completely banned at all the restaurants in the country”, while ICLE covered a wide range of topics, and ICE did not have topic metadata. Nevertheless, all three resources represent the same genre (i.e., student essays), which makes them sufficiently comparable for this study, though potential topic effects should be considered when interpreting our results. We also acknowledge that differences between the native written and spoken data (AmE vs. BrE) might partly explain some of the differences between the two corpora. However, it is challenging to find suitable British student written data or American student spoken data. More importantly, mixing different native varieties (BrE and AmE) within the same corpus would undermine the comparability of our corpora that were designed using the same systematic criteria within each mode. Table 1 provides an overview of the corpora used in this study, along with their word counts and the frequency of TAKE in each corpus.

Table 1. Overview of the corpora.

	Variety	Mode	Corpus	No. of Words	No. of TAKE
LEs	MCE	Written	ICLE-MC + ICNALE-MC-stw	150,130	432
		Spoken	LINDSEI-MC	63,493	87
	BFE	Written	ICLE-FR	194,025	375
		Spoken	LINDSEI-FR	90,999	113
NEs	SgE	Written	ICE-SIN-stw + ICNALE-SIN-stw	143,446	378
		Spoken	NESSI-SIN	136,527	257
	HKE	Written	ICLE-HK	384,016	510
		Spoken	NESSI-HK	113,912	157
NativeE	AmE	Written	LOCNESS	167,385	329
	BrE	Spoken	LOCNEC	122,132	162
Total				1,566,065	2800

Apart from the above corpora, the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) were used in this study. They were not used for direct comparison with non-native corpora but rather for spot checks when studying possible non-standard features found in the target corpora. Valency patterns and senses were classified as non-standard when they did not appear in either the BNC or COCA. For ambiguous instances (e.g., low-frequency or suspicious uses), further evaluation was conducted among native speakers of both AmE and BrE to determine the naturalness of valency patterns and senses.

4.2. Methods

All instances of TAKE were retrieved from the target corpora using the corpus tool LancsBox 6.0 (Brezina et al., 2021). Instances of repetition and incompleteness were manually excluded from the analysis. For repetitions, we removed cases where the verb was immediately repeated without adding new information, as illustrated in (1). For incomplete utterances, we excluded cases where the valency patterns and senses of the target verb could not be determined due to limited contextual information, as shown in (2). In total, 2800 instances of TAKE were kept for further analysis.

- (1) B: I don't know take care of him **take take** him out (NESSI-HK-025)
- (2) A: wha = what what are the differences between your . poly study experience and your university experience
B: I think . for one it was much easier to make friends in poly . because like my cohort in poly was really small . and we **took** like . my . basically I never changed classes the whole way the whole three years that we were there (NESSI-SIN-041)

All instances of TAKE were manually coded for valency patterns and senses with reference to the Valency Dictionary of English (VDE) (Herbst et al., 2004). The VDE was developed within valency theory, whose basic assumption is that the verb occupies a central position in a sentence and determines how other elements combine with it to form a grammatical sentence (Herbst et al., 2004, p. xxiv). This makes valency theory particularly suitable for the investigation of verb-specific features at the lexis-grammar interface across different English varieties. Furthermore, the VDE provides important reference data, enabling systematic coding procedures across all varieties investigated in this study. According to the VDE, the valency patterns of TAKE specify the morphological-formal categories⁵ of each complement (e.g., noun, adjective, and adverbial) (Herbst et al., 2004, p. 843). For instance, the valency pattern of TAKE in (3) can be described as SCU⁶ + TAKE + NP + as-NP.

- (3) They **take** television as an important part in daily life rather than a tool through which they can learn knowledge. (ICLE-CNUK-1122)

It should be noted that all passive sentences were treated like their active counterparts in this study, unless only passive sentences were found in the corpora. This decision was motivated by the observation that active-passive alternations typically preserve the core meaning of the verb (Hanks, 2013).

The senses of TAKE were identified following the VDE (Herbst et al., 2004, p. 846), with concise formal labels⁷ adapted from Gilquin (2008) for analytical consistency (see Table 2).⁸ For instance, TAKE in (3) conveys the meaning of *consider*. Given that this study examines non-native data and the VDE does not provide exhaustive semantic descriptions comparable to traditional dictionaries (Herbst et al., 2004, p. xxxviii), the *other* category was added to record senses not covered in the VDE and instances of ambiguity. To ensure coding reliability, intra-rater reliability (i.e., agreement of a single rater across time) (McHugh, 2012, p. 277) was calculated. To do so, a random sample of 100 concordance

lines with TAKE was re-coded after around two months. Cohen’s kappa test showed that there was a very good agreement between the two rounds ($k > 0.90$, $p < 0.01$).

Table 2. Senses of TAKE.

Senses	Examples
1 Grab	So the artist (er) you know took his paintbrush... (NESSI-HK-009)
2 Move	He was taken to the Tower of London... (LOCNEC-EN043)
3 Habitual actions and qualities	But they are more likely to take a drug... (ICLE-CN HK1361)
4 Require	It takes quite a long time to reach it. (ICLE-FRUC2009)
5 Travel	Do you wanna take a taxi there. (NESSI-SIN-040)
6 Engage in	The manager had an opportunity to take a computer class (ICLE-US-MICH-0002.1)
7 Do (delexical sense)	They should not have to take part in a religion that... (LOCNESS-USMRQ0015.1)
8 Capture	it started with the Romans trying to take Scotland... (LOCNEC- EN043)
9 Consider	I took it as seriously as a film... (NESSI-SIN-021)
10 Assume	I tend to take more leadership roles... (NESSI-SIN-028)
11 Particle verbs	In conclusion, taking up part-time job... (ICNALE-WE_SIN_PTJ0_003_B1_2)
12 Idioms	This action takes place in a number of different countries. (LOCNESS-US-PRB-0023.1)
13 Accept	Another benefit to taking these jobs... (ICNALE-WE_SIN_PTJ0_046_B2_0)
14 Use	The wealthy countries will take an opportunity to help... (ICLE- FRUC1076)
15 Other	More graduate men also took brides of equal qualifications. (ICE-SIN-W1A-003)

For data analysis, hierarchical clustering analysis (HCA) was used to explore the degree of (dis)similarity between varieties regarding the distribution of valency patterns and the senses of TAKE. Following previous studies (e.g., [Gilquin & Granger, 2021](#); [Laporte, 2021](#)), the Euclidean distance metric as a measure of (dis)similarity and Ward’s rule as an amalgamation strategy were employed when conducting HCA. The output of the HCA is a dendrogram with several clusters displaying the distance between varieties. All plots and statistical analyses were generated using R (version 4.5.2; [R Core Team, 2025](#)).

5. Results and Discussion

5.1. Valency Patterns of TAKE Across Varieties and Modes

Figures 2 and 3 visualize the percentages of the top 5 most frequent valency patterns of TAKE in writing and speech in each variety using a heatmap with a dendrogram based on HCA. In each heatmap,⁹ colour intensities (from light to dark) represent the percentages of valency patterns, with darker colours indicating a higher percentage of the pattern. The dendrogram based on HCA at the top illustrates the relationships between the varieties in terms of the distribution of valency patterns. In the dendrogram, the length of the vertical line indicates the distance between clusters. The lower merges suggest greater similarity between items (English varieties here), while higher merges indicate dissimilarity ([Levshina, 2015](#), p. 309). Furthermore, each heatmap includes a distance annotation at the bottom that specifies how far each variety deviates from NativeE based on overall valency distribution. This is visualized through a mint green gradient where lighter shades indicate closer similarity to NativeE (lower distance values) and darker shades represent greater divergence (higher distance values). For further details, including raw frequency, rankings, and percentage distribution of the top 5 valency patterns in writing and speech across English varieties, see Tables A1 and A2 in the Appendix A. Due to space limitations, we only display the top 5 valency patterns, which account for 69.5% (NativeE) to 93.6% (HKE) of all patterns across varieties. Since the top 5 valency patterns differ across varieties, the tables and figures include more than five patterns.

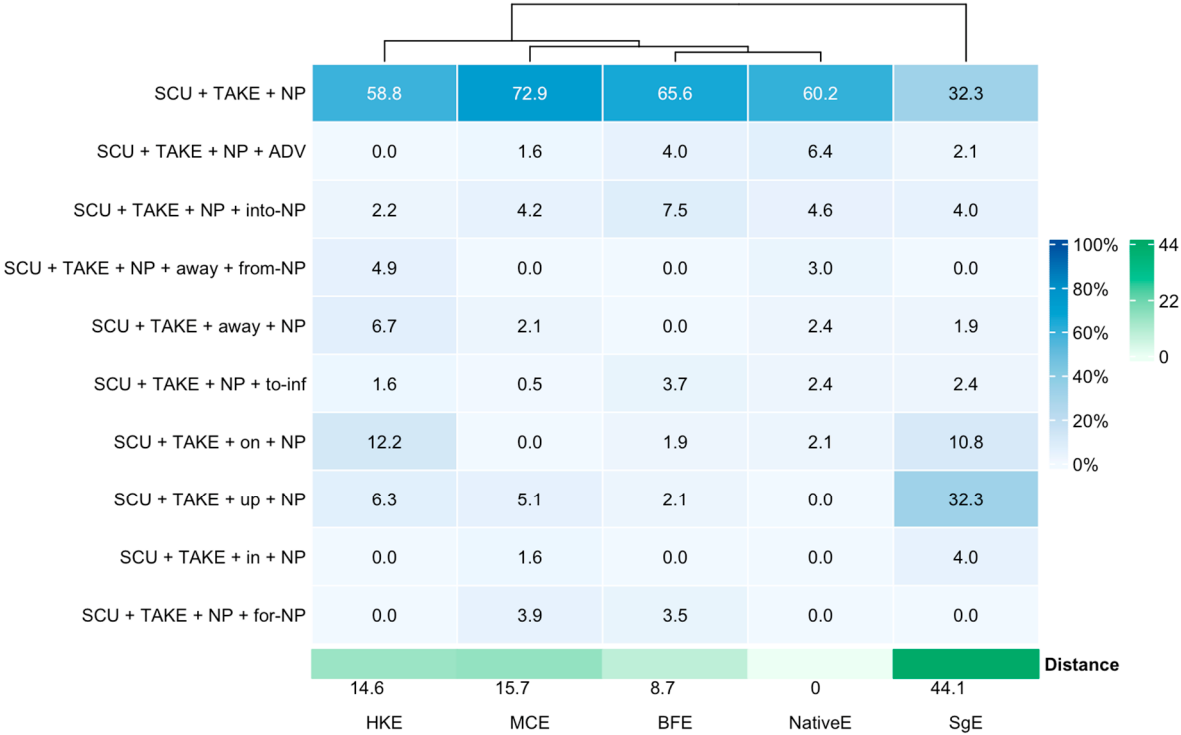


Figure 2. Heatmap of the distribution of the valency patterns of TAKE in writing across varieties (%).

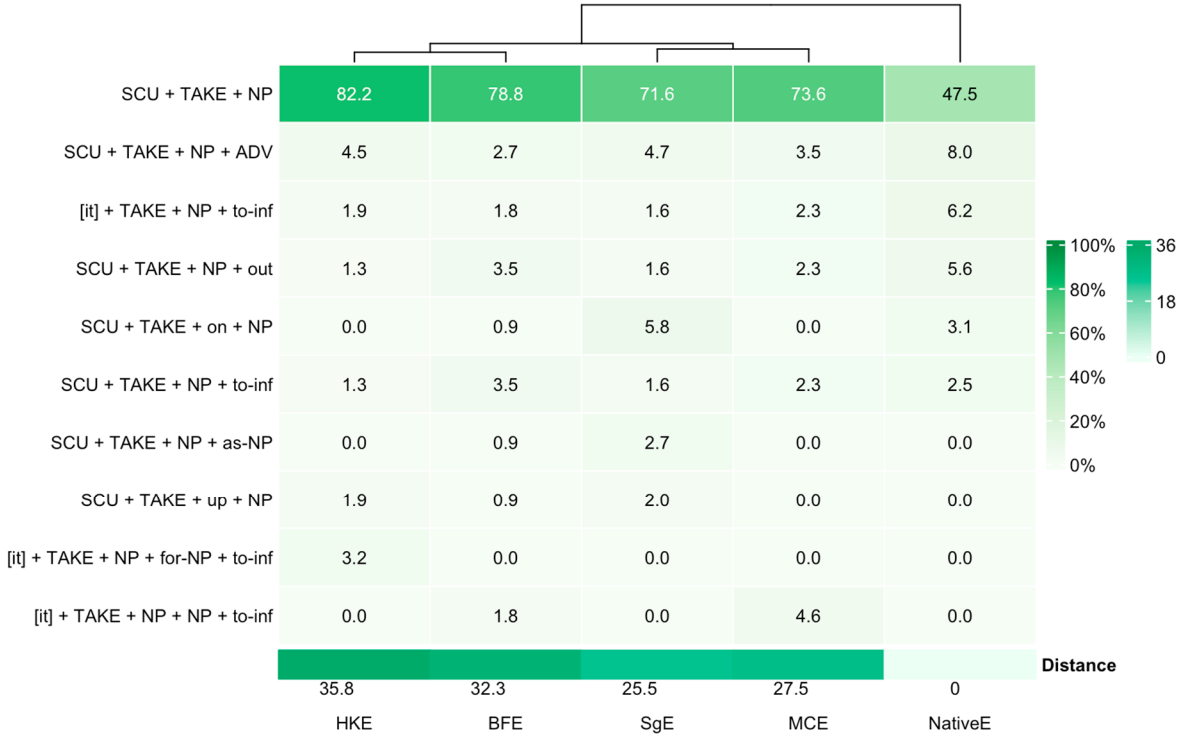


Figure 3. Heatmap of the distribution of the valency patterns of TAKE in speech across varieties (%).

5.1.1. Valency Patterns of TAKE Across Varieties in Written Mode

Several interesting findings emerge from Figure 2, which presents the distribution of the top 5 valency patterns of TAKE in writing across varieties. First, the distribution of valency patterns of TAKE is heavily concentrated in a few core patterns (i.e., the top 5) in writing across varieties. NativeE, however, displays a less concentrated distribution (cumulative percentage of 76.6% for the top 5) than non-native English varieties, among

which SgE exhibits the least concentrated distribution (83.4%), followed by BFE (84.3%), MCE (88.2%), and HKE (88.9%). This suggests that, compared to NativeE writers, both LEs and NEs writers show greater concentration in the top valency patterns of TAKE. This supports broader findings that non-native English varieties often exhibit a more restricted set of syntactic patterns (e.g., Edwards & Laporte, 2015). Yet, the trend is not uniform within the two types of varieties. Both LEs and NEs show similar internal variation, with MCE (88.2%) and HKE (88.9%) exhibiting higher concentration levels and BFE (84.3%) and SgE (83.4%) showing lower concentration levels. The variation within both LEs and NEs groups demonstrates the potential limitation of the aggregate approach and the importance of carefully examining each variety individually (cf. Section 3).

A second finding worth highlighting is that the most frequent valency pattern (i.e., *SCU + TAKE + NP*) is shared across all English varieties, as illustrated in (4) and (5).

- (4) We must now be able to **take** responsibility for this behavior. (LOCNESS-US-PRB-0035.2)
- (5) I think that when you have to **take** decisions, the support of people is needed. (ICLE-FRUL1008)

However, the proportional dominance of this pattern varies across varieties, with 72.9% in MCE, followed by BFE (65.6%), NativeE (60.2%), and HKE (58.8%). Most striking is the markedly lower proportion in SgE (32.3%), due to its reliance on another pattern, *SCU + TAKE + up + NP* (32.3%), which remains very rare or absent among the top 5 valency patterns in the other varieties, ranging between 0% (NativeE) and 6.3% (HKE). This suggests that SgE, as the most advanced variety, may develop distinct pattern preferences from both native and other non-native varieties. A closer look at the concordance lines reveals that this unusually high frequency of *SCU + TAKE + up + NP* in SgE is primarily driven by the recurrent expression *TAKE up (a) part-time job(s)* in ICNALE-SIN (see (6)). This is clearly a strong topical focus in ICNALE, which was compiled based on two specific topics, i.e., *a part-time job* and *smoking* (see Section 4.1). However, in MCE writing, *a part-time job(s)* tends to collocate with TAKE rather than TAKE *up*, reflecting the influence of ICNALE-MC with its specific topical focus, as seen in (7).

- (6) Therefore, it is good that college students **take up** a part-time job, but it is definitely not as important as the focus they should place on their studies. (ICNALE-SIN-WE_SIN_PTJ0_184_B1_2)
- (7) Above all, I think it is important for college students to **take** a part-time job. (ICNALE-MC_CHN_PTJ0_170_B1_2)

Third, the top 5 valency patterns vary significantly across varieties except for *SCU + TAKE + NP*, appearing consistently as the most frequent pattern. This heterogeneous distribution suggests variety-specific preferences for the valency patterns of TAKE. As shown in Figure 2, some of the top 5 valency patterns of TAKE in certain varieties are completely absent in others (see 0% cells). For instance, *SCU + TAKE + in + NP* (see (8)) is the fourth most frequent pattern in SgE (4%) but is absent in the top 5 ranks of NativeE, BFE, and HKE, indicating clear variety-specific preferences.

- (8) ...this way of fragmenting the sentence does not correspond to the way the human brain **takes in** information. (ICE-SIN-W1A-006)

The dendrogram based on HCA for written data in Figure 2 reveals two distinct clusters in the distribution of valency patterns of TAKE across varieties: one comprising HKE, MCE, BFE, and NativeE, and the other consisting of SgE only ({HKE-MCE-BFE-NativeE}-{SgE}). In the first cluster, NativeE and BFE cluster together first because of their more balanced distributions of valency patterns of TAKE. This primary cluster is then joined by MCE and HKE, which exhibit intermediate similarity to NativeE. The separate cluster of

SgE can be explained by its distinctively high proportion of the pattern *SCU + TAKE + up + NP* (32.3%). It is noteworthy that HKE (an NE) does not cluster with SgE (another NE) but with two LEs. This observation supports the previous finding that individual varieties do not necessarily cluster together according to their assigned variety status within the LEs-NEs continuum (e.g., Deshors, 2014; Edwards & Laporte, 2015; Gilquin & Granger, 2021). According to the Euclidean distance, BFE (8.7) shows the greatest similarity to NativeE, followed by HKE (14.6), MCE (15.7), and finally SgE (44.1) (**BFE > HKE > MCE > SgE**), which differs from the expected proximity to NativeE (**BFE > MCE > HKE > SgE**) established within the EIF Model. Specifically, HKE and MCE show reversed positioning compared to the EIF Model's predictions.

5.1.2. Valency Patterns of TAKE Across Varieties in Spoken Mode

Figure 3 presents the distribution of the valency patterns of TAKE in speech across varieties. First of all, it appears that, like the written data, the overall distribution of the valency patterns of TAKE in speech is heavily concentrated in a few patterns across varieties (see also Table A2 in the Appendix A). This concentration is, however, more pronounced in writing than in speech in NativeE, with the top 5 valency patterns of TAKE accounting for 76.6% in writing versus 69.5% in speech. This trend in NativeE might be partly attributed to the use of different native varieties for the written (AmE) and spoken (BrE) control corpora in this study, rather than merely being a mode-related difference. By contrast, the remaining non-native English varieties show the opposite trend: the concentration is more pronounced in speech than in writing, with the top 5 valency patterns in SgE showing a cumulative percentage of 87.2% in speech vs. 83.4% in writing, HKE 93.6% vs. 88.9%, and BFE 90.3% vs. 84.3%. Nevertheless, as in writing, NativeE still displays the least concentrated distribution of all varieties in speech (69.5%). Among non-native English varieties, SgE again exhibits the least concentrated distribution (86.8%), but this time followed by MCE (87.4%), BFE (90.3%), and HKE (93.6%).

A second interesting finding that emerges from Figure 3 is that the most frequent valency pattern is still *SCU + TAKE + NP* across all English varieties in speech. However, this pattern shows notable proportional differences across modes. In NativeE, this pattern is proportionally much less frequent in speech (47.5%) than in writing (60.2%), while non-native varieties show the opposite trend, with proportionally more occurrences in speech than in writing. Particularly striking is the case of SgE, where the distribution in speech becomes more similar to the other non-native English varieties compared to its distinctive distribution in writing. This might be attributed to the spontaneous nature of speech, which often favours shorter, less cognitively demanding patterns (e.g., Miller & Weinert, 1998; Leech, 2000). Furthermore, variety-specific preferences for certain valency patterns of TAKE are also observed in spoken data. Notably, certain patterns appear among the top 5 most frequent patterns in specific varieties while being absent in others (see 0% cells).

Third, different valency patterns of TAKE are preferred in writing vs. speech across English varieties. For instance, patterns with *it*-extraposition (e.g., *[it] + TAKE + NP + to-inf*) occur among the top 5 most frequent patterns in speech, but not in the top 5 ranks in writing. In writing, *SCU + TAKE + NP + to-inf* is usually used alternatively to convey the same meaning. Examples (9) and (10) illustrate this difference. This distinctive pattern highlights the importance of examining both written and spoken data when comparing LEs and NEs.

(9) It **takes** quite a long time to get out (LOCNEC-EN020)

(10) Amy **took** a great deal of time to try and express what she had seen on television.
(LOCNESS-US-MRQ-0034.1)

The dendrogram for spoken data in Figure 3 reveals two distinct clusters that differ from the results based on written data, pointing to the significant impact of mode on the clustering patterns. More specifically, non-native varieties cluster together, while NativeE stands out as a separate cluster ({NativeE}-{HKE-BFE-SgE-MCE}). These clusters probably reflect the distinctive proportions of the most frequent pattern (i.e., *SCU + TAKE + NP*) among varieties. NativeE is distinguished by its lower proportion of the most frequent pattern (47.5%) compared to all non-native English varieties, which show higher proportions, ranging from 71.6% in SgE to 82.2% in HKE. Within the non-native cluster, the groupings cut across variety-type: BFE (LE) groups with HKE (NE), while MCE (LE) groups with SgE (NE). This again suggests that individual varieties do not necessarily cluster together according to their assigned variety status. Most significantly, the spoken data also challenge the expected proximity to NativeE, as SgE (25.5) emerges as the closest variety to NativeE, followed by MCE (27.5), BFE (32.3), and HKE (35.8) (i.e., **SgE > MCE > BFE > HKE**). Importantly, except for SgE (44.1), these distance values are considerably higher than those in written data (15.7 in MCE, 8.7 in BFE, and 14.6 in HKE), indicating that non-native varieties likely show greater overall divergence from NativeE in spoken language. These observations reveal that mode has a significant influence on both the distribution of valency patterns of TAKE across varieties and the positioning of individual varieties along the LEs-NEs continuum. Our results also support Deshors's (2014, p. 298) observation that individual LEs and NEs intermingle within the continuum and position themselves distinctively closer or further away from NativeE. This intermingled clustering, however, cannot be detected in research using the aggregate approach (see Section 3) and again highlights the importance of investigating individual varieties when comparing LEs and NEs (see Laporte, 2012).

5.2. Senses of TAKE Across Varieties and Modes

To explore the potential effect of different linguistic levels on the positioning of the individual LEs and NEs along the continuum, we now investigate another linguistic level, namely, the senses of TAKE. This analysis allows us to see whether different linguistic levels reveal distinct clustering patterns and proximity sequences. As with the analysis of valency patterns, Figures 4 and 5 visualize the semantic distribution of TAKE using heatmaps with dendrograms, with distance annotation indicating the closeness between NativeE and each of the other varieties. For further details, including raw frequency, rankings, and percentage distribution of the senses of TAKE in writing and speech across English varieties, see Tables A3 and A4 in the Appendix A.

5.2.1. Senses of TAKE Across Varieties in Written Mode

Figure 4 presents the semantic distribution of TAKE in writing across varieties. Most notably, like the valency patterns of TAKE, the semantic distribution of TAKE is concentrated towards a few senses across varieties, with the top 5 senses representing 74.5% in NativeE, 89.2% in SgE, 86.3% in HKE, 83.1% in MCE, and 82.2% in BFE (see Table A3 in the Appendix A). This indicates that both LEs and NEs show greater concentration in the top senses of TAKE than NativeE. Importantly, different from the finding on valency patterns in writing, where the trend was not uniform within the two types of varieties, LEs and NEs display a consistent concentration trend this time, with a higher concentration level in NEs than LEs.

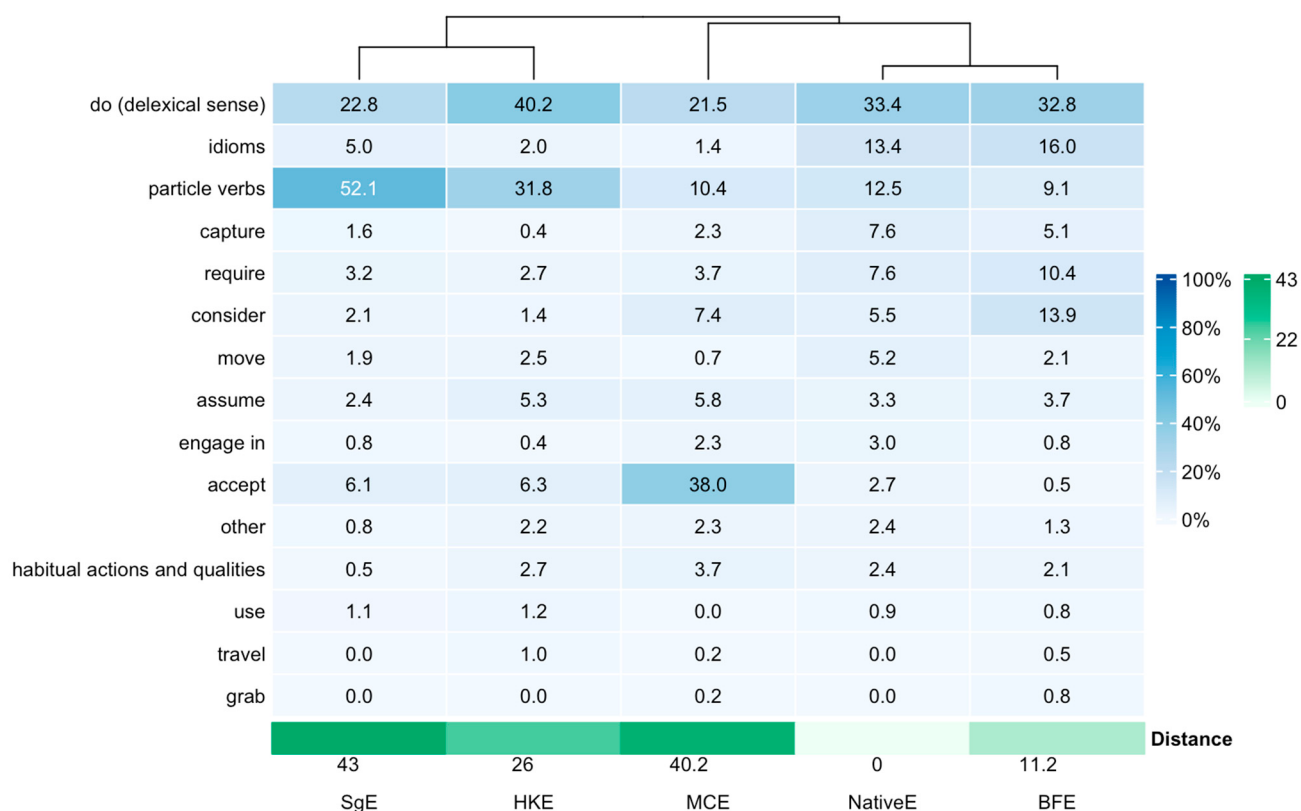


Figure 4. Heatmap of the semantic distribution of TAKE in writing across varieties (%).

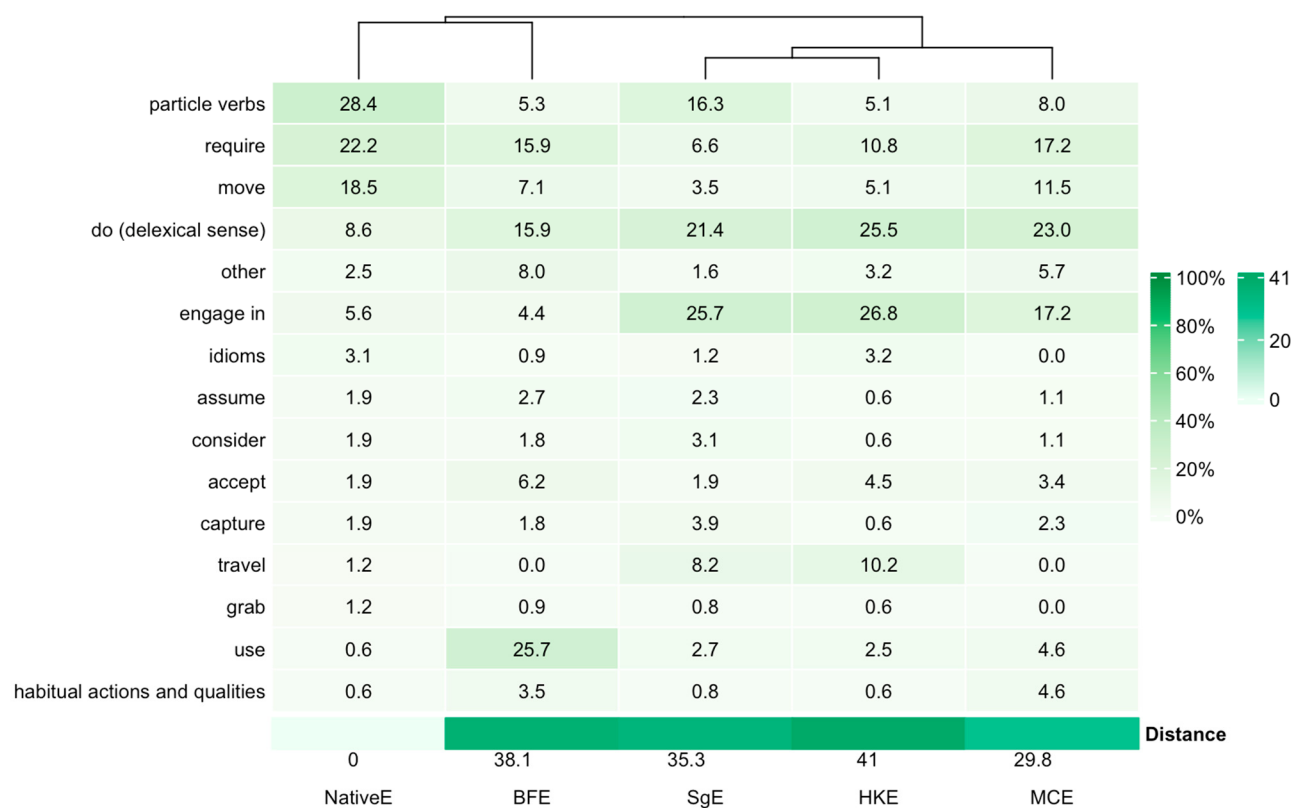


Figure 5. Heatmap of the semantic distribution of TAKE in speech across varieties (%).

Another striking finding concerns the most frequent sense of TAKE in writing across varieties. *Do* (delexical sense) (11) predominates in NativeE (33.4%), HKE (40.2%), and BFE

(32.8%), whereas *particle verbs* (12) are the most common sense in SgE (52.1%), and *accept* prevails in MCE (38.0%). The predominance of delexical TAKE in NativeE confirms previous findings (Gilquin, 2008), while the heterogeneous picture in non-native varieties supports Werner and Mukherjee's (2012) observation regarding NEs (Indian English and Sri Lankan English). The preference for *particle verbs* in SgE and *accept* in MCE likely reflects the earlier discussion on the potential topic effect in the ICNALE, where SgE and MCE exhibit different collocational preferences when talking about *having a part-time job(s)*: SgE favours *take up a part-time job(s)* (i.e., *particle verbs*) while MCE favours *take a part-time job(s)* (i.e., *accept*). This observation demonstrates how the topic can influence both the valency patterns and the semantic distribution of lexical items. In addition, different varieties display different preferences for the remaining top senses. For instance, *consider* ranks among the top 5 senses in MCE and BFE, but not in the other varieties, indicating variety-specific semantic preferences.

- (11) He or she has **taken the option** of a non-marketplace, non-public and non-financially rewarded job. (LOCNESS-US-IND-0004.1)
- (12) However, having a co-curricular activity or **taking up** a leadership position in various groups can better enhance the students' college life. (ICNALE-WE_SIN_PTJ0_004_B2_0)

Finally, among the different varieties, the *other* category represents the highest proportion in NativeE (2.4%), followed by MCE (2.3%), HKE (2.2%), BFE (1.3%), and SgE (0.8%). As a reminder, the *other* category refers to cases where TAKE cannot be assigned to senses listed in the VDE, as seen in (13) and (14).

- (13) Their struggles and hopes of forty years will not have been in vain—Without the “events of Berlin Wall” history probably would not have **taken** a very different course. (LOCNESS-US-MICH-0024.1) (sense: *follow*)
- (14) Here, I would like to recommend students to **take** a good habit to have a record for the use of credit card. (ICLE-CN HK1402) (sense: *develop*)

A similarly higher proportion of the *other* category of TAKE in NativeE than in NEs was found in Werner and Mukherjee (2012),¹⁰ who attributed this phenomenon to the more advanced status of NativeE. That is to say, the semantic extension process is arguably linked to varieties' developmental phases. However, our findings suggest a more complex picture: SgE, the most advanced NE, exhibits the lowest degree of semantic extension, challenging the linear evolutionary model. We will return to the *other* category in Section 5.3.2.

The dendrogram based on the senses of TAKE in the written mode shows different clustering compared to the results based on the valency patterns, indicating the effect of linguistic levels on the clustering pattern. NativeE, BFE, and MCE form the first cluster, while SgE and HKE constitute the second cluster ({SgE-HKE}-{MCE-NativeE-BFE}). The separate cluster of SgE and HKE could be explained by their unusually high proportions of *particle verbs* (52.1% and 31.8%). Notably, it is the first time that we observe individual varieties clustering together according to their assigned variety status (NEs vs. LEs and NativeE). This result differs from our earlier clustering for the valency patterns of TAKE and the heterogeneity reported in previous studies on the semantic distribution of MAKE (Laporte, 2012) and into (Edwards & Laporte, 2015). This is different from the clusters based on valency patterns in writing (cf. Figure 2), where HKE clusters with the MCE and BFE (two LEs) instead of SgE (another NE). This inconsistency seems to support Gilquin and Meriläinen's (2024) suggestion that different linguistic levels may lead to different conclusions regarding the relationship between LEs and NEs (two distinct categories or a continuum). However, the Euclidean distance values demonstrate the same proximity to NativeE as the results for valency patterns in writing: BFE (11.2) > HKE (26) > MCE (40.2) > SgE (43).

5.2.2. Senses of TAKE Across Varieties in Spoken Mode

Figure 5 presents the semantic distribution of TAKE in speech across varieties. Similar to written data, the semantic distribution of TAKE in speech is concentrated towards a few senses across varieties, with the top 5 senses accounting for 83.3% in NativeE, 78.2% in SgE, 78.4% in HKE, 76.9% in MCE, and 72.6% in BFE. This concentration is, however, more pronounced in speech than in writing in NativeE, with the top 5 senses of TAKE accounting for 77.1% in speech versus 74.5% in writing, respectively. In contrast, the remaining non-native English varieties show the opposite trend, with the concentration being more pronounced in writing than in speech: 89.2% vs. 78.2% for SgE, 86.3% vs. 78.4% for HKE, 83.1% vs. 76.9% for MCE, and 82.2% vs. 72.6% for BFE.

Furthermore, it turns out that the most frequent sense of TAKE in speech varies across varieties. *Particle verbs* are the most frequent category in NativeE (28.4%), *engage in* predominates in SgE (25.7%) and HKE (26.8%), *do* (*delexical sense*) prevails in MCE (23.0%), and *use* is the most common sense in BFE (25.7%). A comparison of Figures 4 and 5 shows that *particle verbs* are more typical of native speech than of native writing. However, all non-native Englishes display the opposite preference, using more *particle verbs* in writing than in speech. This suggests that both LEs and NEs speakers have little awareness of mode differences in NativeE, which differs from Gilquin's (2011, 2016a) findings that NEs speakers are better than LEs speakers at reproducing the native distribution of *particle verbs* across modes. Similarly, varieties display different preferences for the remaining top senses. For instance, *engage in* is among the top 5 senses in NativeE, SgE, HKE, and MCE, but not in BFE.

Turning to the *other* category, a striking contrast emerges between spoken and written modes. While written data partially support Werner and Mukherjee's (2012) finding that more advanced varieties exhibit greater semantic extension, with NativeE showing the highest proportion of the *other* category, the spoken data display quite a different trend: BFE (the least advanced variety) accounts for the largest proportion (8.0%), followed by MCE (5.7%), HKE (3.2%), NativeE (2.5%), and SgE (1.6%). The trend in spoken data suggests an inverse relationship between the semantic extension and the developmental phases of English varieties: the more advanced the variety, the lower the proportion of the *other* category. This relationship is particularly evident among non-native varieties, though NativeE (the most advanced variety) shows a slightly higher proportion than SgE.

The dendrogram in Figure 5 reveals a different clustering pattern from the written data. It is still formed by two main branches, but they cut across varietal types. On one branch, NativeE and BFE still cluster together. However, on the other branch, SgE and HKE (NEs) cluster with MCE (LE) ({NativeE-BFE}-{SgE-HKE-MCE}). This clustering seems to suggest that varieties with similar mother tongue backgrounds (i.e., SgE, HKE, and MCE) might retain similar usage patterns, pointing to L1 transfer as a potential factor in the semantic distribution of TAKE. As for the cline of proximity to NativeE, the Euclidean distance values based on the semantic distribution of TAKE in speech show that MCE (29.8) is closest to NativeE this time, followed by SgE (35.3), BFE (38.1), and HKE (41) (MCE > SgE > BFE > HKE). This cline, once again, does not reflect the expected proximity to NativeE established within the EIF Model.

The comparison between written and spoken data demonstrates that mode has a significant influence on the semantic distribution of TAKE across varieties, thus affecting inter-variety similarities. This divergence, again, underscores the importance of examining both writing and speech when comparing LEs and NEs. Compared to the clusters based on valency patterns in the spoken mode, the clusters here also illustrate the effect of linguistic levels on the positioning of individual English varieties along the continuum.

The above quantitative analysis has uncovered the distributional features of valency patterns and senses of TAKE across varieties and modes. On that basis, we find that individual varieties are intermingled along the LEs-NEs continuum rather than occupying a stable position according to their assigned variety status, with proximity to NativeE varying depending on the mode and linguistic levels investigated. However, such results do not reveal whether there are new valency patterns and senses that are different from standard usages (i.e., non-standard features) in LEs and NEs. The following section thus presents detailed qualitative analyses of selected non-standard valency patterns and senses of TAKE that are common in LEs and NEs and examines the motivations that drive these non-standard features.

5.3. Non-Standard Features of Valency Patterns and Senses of TAKE Across LEs and NEs

According to Croft (2000, p. 8), the emergence of non-standard uses involves the remapping between form and meaning (e.g., grammatical structure and semantics/pragmatics) and is driven by various internal and external linguistic mechanisms, including cognitive processes, language-internal irregularities, and language contact and transfer (Deshors et al., 2018, p. 8). These mechanisms and remapping phenomena are common across all English varieties, allowing for meaningful comparisons between LEs and NEs at this level (Van Rooy, 2011; Meriläinen, 2017).

5.3.1. Non-Standard Features of Valency Patterns of TAKE Across LEs and NEs

The analysis reveals two main categories of non-standard features of valency patterns: (1) lexical non-standard features, which involve minor lexical modifications, particularly in preposition uses (e.g., *take into consideration of*), and (2) syntactic non-standard features, which concern syntactic alternations (e.g., *to-inf* vs. *-ing*) in a valency pattern. These non-standard valency patterns manifest new pattern-existing meaning remapping.

Lexical Non-Standard Features

In some valency patterns, the realization of certain slots allows for lexical variants that deviate from standard English. One such variation pertains to prepositional complements, particularly the insertion or misuse of prepositions. Prepositions are often regarded as the *bête noire* for both teachers and learners because they bear little meaning, making them challenging to learn and teach (Gilquin & Granger, 2011). As a result, they may lead to some variation in non-native English varieties (Gilquin & Granger, 2011; Edwards & Laporte, 2015). For instance, (15) and (16) illustrate the insertion of the preposition *of* in the pattern SCU + TAKE + *into-NP* + *of-NP* in SgE and MCE.

- (15) Laws should not be of a majoritarian model but **take into consideration of** the minority as well. (ICNALE-SIN-WE_SIN_SMK0_071_B2_0)
- (16) **Taking into account of** all these factors, we can safely draw a conclusion that university degrees are theoretical but necessary, nowadays. (ICLE-MC-CNUK1158)

Edwards and Laporte (2015, p. 159) pointed out that the formation of the pattern SCU + TAKE + *into-NP* + *of-NP* in their HKE data (ICE-HK) could be explained by analogy to existing patterns, i.e., *take into account* and *take account of*. Since *take into consideration of* is largely synonymous with the former two expressions, one can further argue that *take into consideration of* is an analogy to *take into account of*. In addition, the superfluous *of* in the above patterns can be explained through analogy with semantically similar patterns such as *think of*. This suggests that analogical processes in the formation of features may operate through multiple and interdependent pathways rather than in isolation (Nesselhauf, 2009).

Another interesting example related to the *take into consideration* pattern is *take in consideration* (or *take in account*), which can be described as SCU + TAKE + *in-NP* + NP. Due

to the obvious semantic connection between *into* and *in*, it is quite common for non-native speakers to confuse these two prepositions (Gilquin & Granger, 2011; Edwards & Laporte, 2015). (17) and (18) illustrate this. It is interesting to note that even native (AmE) speakers (18) seem confused about these two prepositions, indicating that the misuse of *in* and *into* may be influenced by language-internal irregularities.

- (17) In the following essay, the pros and cons of abortion will be discussed and the adoption of abortion will be **taken in consideration**. (ICLE-HK-CN HK1772)
- (18) However, when **taken in account** the effects of peer pressure, honor codes are weakened. (LOCNESS-US-MRQ-0044.1)

Regarding minor lexical modifications, the adverbial particle *away* is found to be inserted in the valency pattern of TAKE in BFE (i.e., SCU + TAKE + NP + *away* + to-NP), as in (19). This pattern may be largely attributed to L1 transfer, as the literal translation of *take me away to restaurants* in French is *m'emmener au restaurant* where *emmener* literally means *take away* in English.

- (19) They: used to **take me: away t= (er) to** restaurants or: to parties and . (em) but the first time I went was when I was sixteen (LINDSEI-FR004)

A second type of lexical non-standard features involves a mix of word classes in the valency patterns of TAKE. For example, writers use the adjectival forms *precedent*, *responsible*, and *careful* instead of the corresponding nominal forms *precedence*, *responsibility*, and *care* in relevant valency patterns of TAKE, including SCU + TAKE + ADJ + *over-NP*, SCU + TAKE + *up* + ADJ, and SCU + TAKE + ADJ, as illustrated in (20) to (22). It is interesting to note that this phenomenon only appears in SgE, HKE, and MCE but not in BFE. This suggests that L1 transfer might be at play in shaping these patterns, since Chinese is an isolating language, which does not distinguish various word classes by morphological changes (Zhao & Jiang, 2020). Example (22) illustrates the confusion between adjective and noun, as both *responsible* and *responsibility* can be translated as *zeren* (literally *responsibility*) in Chinese. This L1 influence might lead Chinese English learners to apply this rule in their English writing.

- (20) In other words, college is the best time for them to fully live out their lives, before adult responsibilities set in and other priorities **take precedent** over their personal desires and needs. (ICNALE-SIN-WE_SIN_PTJ0_044_B2_0)
- (21) They can't neglect the right of the babies. If they don't want to have a baby parents should **take careful** of their behaviours. Abortion is not a way to solve the problem. (ICLE-HK-CN HK1322)
- (22) Some of the criminals are not mean to kill other people and they do want to have an opportunity to confess while **take responsible** of hurting others. (ICLE-MCE-CNUK1150)

Syntactic Non-Standard Features

The second category of non-standard features pertains to syntactic variation in complement patterns, such as the competition between gerunds (*-ing*) and infinitives (*to-inf*), which has received some attention in non-native Englishes (e.g., Deshors & Gries, 2016; Romasanta, 2020; Laporte, 2021). In (23) and (24), the valency pattern of TAKE can be described as SCU + TAKE + NP + *inf*, with a bare infinitive instead of the more standard *V-ing* or *to-inf*.

- (23) We have a system where we each **take turns cook**. (LOCNESS-US-MICH-0043.1)
- (24) I was like this is what we're going to talk about let's try to be civil about everybody **take turns hear** each other out. (NESSI-SIN-033)

Examples (25) to (27) illustrate the non-standard valency patterns *SCU + TAKE + NP + V-ing* and *SCU + TAKE + NP + NP + V-ing* in NativeE, BFE, and MCE, with *to-inf* being used instead of the *V-ing* form.

- (25) (mm) .. cos I know in Germany like you d = you can **take years getting** a degree because you can just keep going (LOCNEC-EN039)
- (26) To make Europe a nation is like to reassemble each piece of a jigsaw: it **takes time doing** it. (ICLE-FRUC1067)
- (27) It **takes us much time working** thus reduces our time for studying. (ICNALE-MC-W_CHN_PTJ0_295_B1_2)

5.3.2. Non-Standard Features of Senses of TAKE Across LEs and NEs

This section examines the non-standard senses of TAKE across LEs and NEs by focusing on the *other* category, which represents cases that cannot fit into established semantic categories of TAKE in the VDE and may contain certain non-standard senses in LEs and NEs contexts. Examination of these cases reveals three potential motivations, namely, metaphorical extension, L1 transfer, and overgeneralization, which may contribute to the formation of non-standard features of TAKE in both NEs and LEs. These non-standard senses illustrate existing form-new meaning remapping.

Metaphorical Extension

The prototypical meaning of TAKE can be described as physical grasping with a movement to the agent (see [Herbst et al., 2004](#), p. 846; [Gilquin, 2008](#), p. 6), as exemplified in *He takes the book from Mary*. This prototypical meaning may provide the cognitive foundation for its semantic extensions across English varieties. For instance, in (28), *European colour* (a cultural identity trait) could be conceptualized as an object that can be grasped and possessed in BFE.

- (28) In other words, their identity will change a bit and **take a European colour**, but a loss of it is hardly conceivable. (ICLE-FRUC1075)

In (29), the *bride* is likely to be metaphorically constructed as an object that can be grasped by cognitive mapping in SgE: obtaining a spouse is grasping an object. It should be noted, however, that this use may also stem from analogy with the established expression *take a wife*.

- (29) More graduate men also **took brides** of equal qualifications. (ICE-SIN-W1A-003)

Similarly, *horror* in (30) seems to be metaphorically conceptualized as a heavy object that cannot be grasped mentally in HKE. However, this could also represent a conventional phrase in standard English, *I can't take it* (meaning: *I can't tolerate it*). Examples (29) and (30) manifest that various internal and external linguistic mechanisms may interact with each other to motivate non-standard features in non-native English varieties (see [Nesselhauf, 2009](#); [Deshors et al., 2018](#)).

- (30) I like comedies sometimes romantic movies (er) sometimes animations but one thing that I definitely do not like is horror or like bloody all the splatter (eh) I just c= I just can't **take it** (NESSI-HK-046)

L1 Transfer

The second motivation for non-standard senses of TAKE is L1 transfer. The L1 equivalents of TAKE is 拿/取 (na/qu) in Chinese ([Wang, 2016](#)) and PRENDRE in French. TAKE *evidence* and *high scores* in (31) and (32) probably stem from direct translations from Chinese: *quzheng* (literally *taking evidence*) and *qude gaofen* (literally *taking a high score*). Similarly, (33) shows possible L1 transfer from French, since *take a good time* is a direct translation

of *prendre du bon temps*. Interestingly, this BFE speaker immediately self-corrects by using *have a good time*, which is more natural in standard English.

- (31) The mistakes in **taking evidence**, disproportionate treatment on the poor and minorities, corruptions in government and many other reasons will lead to the result of taking an innocent person's life. (ICLE-CNUK1150)
- (32) **Taking high score** is the most important of all for a student in China because this is the warranty that others judge whether you are good or not. (ICLE-CNUK1183)
- (33) yes . but now I'm I can really (er) . enjoy my living here and I can (er) **take a goo**= I can have a good time in Louvain-la-Neuve and (er) (LINDSEI-FR-050)

Overgeneralization

The final possible motivation for non-standard senses of TAKE is overgeneralization. Examples (34) to (40) illustrate this, with speakers who tend to use TAKE as a default option rather than using more precise verbs. This observation supports the so-called “lexical teddy bear” effect (Hasselgren, 1994), a cognitive avoidance strategy where learners rely on familiar high-frequency verbs when they have difficulty in finding more specific verbs. Similar tendencies have been documented across non-native English varieties with other highly frequent verbs (e.g., Liu & Shaw, 2001; Laporte, 2012; Wang, 2016). It is also interesting to see that TAKE is used instead of other highly frequent verbs in our examples (see (37) to (40)). While highly frequent verbs are shown to sometimes collocate with identical nouns in delexical verb constructions, they display semantic and syntactic differences (see Dixon, 2005; Giparaitė, 2016). For instance, TAKE *a look* involves some effort while GIVE *a look* expresses a short look affecting the object in some way (Dixon, 2005, pp. 473–474). The substitution phenomena observed in our data seem to indicate that highly frequent verbs like TAKE, HAVE, GIVE, and MAKE are undergoing delexicalization processes in other structures in LEs and NEs. More specifically, these highly frequent verbs are becoming so semantically bleached that they are more interchangeable for non-native English speakers. Importantly, the overgeneralization of TAKE in our examples appears to reflect how the “lexical teddy bear” effect can override linguistic accuracy, creating fossilized interlanguage features that resist native-like development. This indicates that SLA phenomena, including language learning strategies (e.g., overgeneralization), interlanguage features, and fossilization, may also operate in NEs contexts.

- (34) Here, I would like to recommend students to **take a good habit** to have a record for the use of credit card. (→ develop) (ICLE-CN HK1402)
- (35) They also **take the bills** by their credit cards. (→ pay) (ICLE-CN HK1572)
- (36) Having a part-time job can **take us more advantages** than disadvantages. (→ bring) (ICLE-MC-W_CHN_PTJ0_234_B1_2)
- (37) I'll **take it into practice** in my college years. (→ put) (ICNALE-MC-W_CHN_PTJ0_180_B1_2)
- (38) It may **take a longterm effect** to customers. (→ have) (ICLE-CN HK1380)
- (39) An illustration may **take this point clear**. (→ make) (ICLE-CNUK1143)
- (40) They have to **take a new lease** of life. (→ give) (ICLE-FRUC3087)

The above qualitative analyses illustrate shared non-standard usages across LEs and NEs, which are driven by a similar set of language learning strategies or cognitive mechanisms, indicating that form-meaning remapping processes operate in both acquisition contexts.

6. Conclusions

This article has examined the effects of language-internal factors (mode and different linguistic levels) on the positioning of two LEs (MCE and BFE) and two NEs (SgE and HKE) along the LEs-NEs continuum and their proximity to NativeE by investigating the valency

patterns and senses of TAKE across written and spoken modes. The results (see Table 3) make it possible to reassess the macro-level LEs-NEs continuum and refine the EIF Model.

Table 3. Variety clusters and proximity to NativeE.

	Written	Spoken
Expected	{BFE > MCE} > {HKE > SgE}	
Valency patterns	{HKE-MCE-BFE-NativeE}-{SgE} BFE > HKE > MCE > SgE	{NativeE}-{HKE-BFE-SgE-MCE} SgE > MCE > BFE > HKE
Senses	{SgE-HKE}-{MCE-NativeE-BFE} BFE > HKE > MCE > SgE	{NativeE-BFE}-{SgE-HKE-MCE} MCE > SgE > BFE > HKE

Note: {} represents the clusters of varieties based on clustering analysis.

For the first two research questions, which focused on mode differences and different linguistic levels in variety positioning along the continuum and their relationship to the expected proximity cline to NativeE (BFE > MCE > HKE > SgE), our quantitative analysis using HCA has revealed that both mode and linguistic levels have a significant influence on the variety positioning along the continuum and its proximity to NativeE. For example, focusing on valency patterns, HKE does not cluster with SgE (another NE) but with the two LEs in written data, while all LEs and NEs cluster together in spoken data. Similarly, the proximity to NativeE changes when we move from the written mode (BFE > HKE > MCE > SgE) to the spoken mode (SgE > MCE > BFE > HKE). When focusing on the spoken mode, we can also observe that the clustering and proximity to NativeE are different at the level of valency patterns ({NativeE}-{HKE-BFE-SgE-MCE} and SgE > MCE > BFE > HKE) and senses ({NativeE-BFE}-{SgE-HKE-MCE} and MCE > SgE > BFE > HKE). These results suggest that LEs and NEs intermingle along the continuum depending on the mode and linguistic levels investigated, rather than occupying a stable position according to their assigned variety status. It shall be noted that the senses of TAKE in the written mode are the only dimension showing the expected LEs/NEs clusters. The other three dimensions, by contrast, support the intermingled clustering. Furthermore, the observed proximity to NativeE consistently deviates from the expected cline across all four dimensions.

For the third research question, which dealt with non-standard features of the valency patterns and senses of TAKE, our qualitative analysis has shown that non-standard features by form-meaning remapping in both LEs and NEs seem to be driven by similar underlying mechanisms, including L1 transfer, analogy, and overgeneralization. This suggests that these non-standard features do not develop at random but follow systematic trends, pointing to the creative ability of both LEs and NEs speakers (Deshors et al., 2018). This provides additional evidence for the inappropriacy of the distinction between “errors” in LEs and “innovations” in NEs (see Section 1).

While several studies (e.g., Deshors, 2014; Edwards & Lange, 2016; Gilquin & Granger, 2021) have established that individual varieties are intermingled rather than grouped according to their assigned variety status, this study further reveals that intermingling clustering is probably constrained by language-internal factors rather than being arbitrary, as demonstrated by the systematic variation in variety clustering across mode (writing vs. speech) and different linguistic levels (valency patterns vs. senses). In light of this, our study reconceptualizes the traditional macro-level LEs-NEs continuum as a dynamic and complex spectrum along which individual varieties exhibit variable development trajectories depending on language-internal factors (e.g., mode and linguistic levels in this study) in addition to the assigned variety status.

Our findings address a critical gap in the current EIF Model. While previous research has highlighted the role of external linguistic factors in variety positioning (cf. Section 2), our study provides a first investigation of how language-internal factors influence the

positioning of varieties along the LEs-NEs continuum. These results demonstrate that language-internal factors, such as mode and linguistic levels, might override or interact with developmental phases in determining variety proximity to NativeE. These findings also have some implications for dealing with the theoretical tension in the field (Section 2). The exposure-based hypothesis predicts that NEs should be closer to NativeE due to greater exposure, while the EIF Model predicts an inverse relationship between developmental phases and proximity to native norms. Our results suggest that both predictions may be partially correct, and their validity depends on specific linguistic conditions. For instance, the exposure-based hypothesis appears valid for valency patterns in the spoken mode, where SgE shows the greatest proximity to NativeE compared to LEs, but the greatest distance of HKE from NativeE contradicts the hypothesis. The EIF Model prediction receives some support from senses in the written mode, where LEs are closer to NativeE than SgE is, but the positioning of HKE, again, complicates the prediction.

Based on these findings, we propose refining the EIF Model (Buschfeld & Kautzsch, 2017; Buschfeld et al., 2018) by incorporating language-internal factors as a fourth dimension. This could supplement the existing third dimension of variety-internal heterogeneity in the current EIF Model (Buschfeld et al., 2018). While the current EIF Model mainly focuses on language-external sociolinguistic factors (e.g., age and social status), our refined model provides a more nuanced theoretical framework to understand the positioning of English varieties. Rather than viewing varieties as occupying fixed positions along the LEs-NEs continuum, which are constrained by developmental phase or exposure to English alone, the refined model acknowledges that all varieties have arguably developed through complex interactions between extra- and intra-territorial forces, as well as language-external and -internal factors.

This refinement calls for a fundamental shift towards multidimensional analyses of variety positioning. A potential implication involves examining systematic correlations between language-internal factors (e.g., mode and linguistic levels) and external sociolinguistic variables (e.g., age, social status, and educational background). For instance, one may wonder whether writing by highly educated groups within a community shows greater proximity to native norms, or whether certain patterns or senses appear more frequently in the speech of the younger generation. Such questions are crucial for determining whether variety positioning is shaped by language-internal factors, language-external factors, or their interaction. Methodologically, this requires rigorous research designs featuring balanced cross-mode datasets, multi-level linguistic analysis, and combining both internal and external variables. Only by doing so can we move beyond static variety categorization towards a dynamic, empirically grounded understanding of English variety development.

Despite the contributions of this study, several limitations must be recognized. First, the written data were collected from different sources with varying topics, which may have potential effects on the results. For instance, the topic about *having a part-time job* in ICNALE led to a higher proportion of *SCU + TAKE + up + NP* in written SgE. Second, the control corpora represent different native English varieties (AmE for writing, BrE for speech) and may thus account for some differences between the written and spoken modes. Third, focusing only on one verb (TAKE) at two linguistic levels (valency patterns and senses) in four varieties (mainly from Asian territories) limits the generalizability of our findings. Future research, therefore, should incorporate more varieties in other regions (e.g., Africa) and investigate additional linguistic phenomena (e.g., from the domains of phonology and pragmatics). In addition, it would be interesting to explore more language-internal factors, such as genre effects and text types, to test the validity of our dynamic LEs-NEs continuum.

Author Contributions: Conceptualization, Y.T. and G.G.; methodology, Y.T. and G.G.; software, Y.T.; validation, Y.T. and G.G.; formal analysis, Y.T. and G.G.; investigation, Y.T. and G.G.; resources, Y.T. and G.G.; data curation, Y.T.; writing—original draft preparation, Y.T.; writing—review and editing, G.G.; visualization, Y.T.; supervision, G.G.; project administration, G.G.; funding acquisition, Y.T. and G.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the China Scholarship Council, grant number 202106730096.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The NESSI data are not publicly available at this stage. The data extracted from the corpus and presented in this study are available on request from the authors.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Note: (#1–#5) indicate ranking within the top 5 patterns or senses in each variety in this and subsequent tables.

Table A1. The top 5 most frequent valency patterns of TAKE in writing in each variety (raw frequencies, rankings, and percentages).

	NativeE	BFE	MCE	HKE	SgE
SCU + TAKE + NP	198 (#1) (60.2%)	246 (#1) (65.6%)	315 (#1) (72.9%)	300 (#1) (58.8%)	122 (#1) (32.3%)
SCU + TAKE + NP + ADV	21 (#2) (6.4%)	15 (#3) (4.0%)	5 (1.6%)	0 (0%)	8 (2.1%)
SCU + TAKE + NP + into-NP	15 (#3) (4.6%)	28 (#2) (7.5%)	18 (#3) (4.2%)	11 (2.2%)	15 (#4) (4.0%)
SCU + TAKE + NP + away + from-NP	10 (#4) (3.0%)	0 (0%)	0 (0%)	25 (#5) (4.9%)	0 (0%)
SCU + TAKE + away + NP	8 (#5) (2.4%)	0 (0%)	9 (#5) (2.1%)	34 (#3) (6.7%)	7 (1.9%)
SCU + TAKE + NP + to-inf	8 (2.4%)	14 (#4) (3.7%)	2 (0.5%)	8 (1.6%)	9 (2.4%)
SCU + TAKE + on + NP	7 (2.1%)	7 (1.9%)	0 (0%)	62 (#2) (12.2%)	41 (#3) (10.8%)
SCU + TAKE + up + NP	0 (0%)	8 (2.1%)	22 (#2) (5.1%)	32 (#4) (6.3%)	122 (#1) (32.3%)
SCU + TAKE + in + NP	0 (0%)	0 (0%)	7 (1.6%)	0 (0%)	15 (#4) (4.0%)
SCU + TAKE + NP + for-NP	0 (0%)	13 (#5) (3.5%)	17 (#4) (3.9%)	0 (0%)	0 (0%)
Cumulative % (top 5)	76.6%	84.3%	88.2%	88.9%	83.4%

Table A2. The top 5 valency patterns of TAKE in speech in each variety (raw frequencies, rankings, and percentages).

	NativeE	BFE	MCE	HKE	SgE
SCU + TAKE + NP	77 (#1) (47.5%)	89 (#1) (78.8%)	64 (#1) (73.6%)	129 (#1) (82.2%)	184 (#1) (71.6%)
SCU + TAKE + NP + ADV	13 (#2) (8.0%)	3 (#3) (2.7%)	3 (#3) (3.45%)	7 (#2) (4.5%)	12 (#3) (4.7%)
[it] + TAKE + NP + to-inf	10 (#3) (6.2%)	2 (#5) (1.8%)	2 (#5) (2.3%)	3 (#4) (1.9%)	4 (1.6%)

Table A2. *Cont.*

	NativeE	BFE	MCE	HKE	SgE
SCU + TAKE + NP + out	9 (#4) (5.6%)	4 (#2) (3.5%)	2 (2.3%)	2 (1.3%)	4 (1.6%)
SCU + TAKE + on + NP	5 (#5) (3.1%)	1 (0.9%)	0 (0%)	0 (0%)	15 (#2) (5.8%)
SCU + TAKE + NP + to-inf	4 (2.5%)	4 (#2) (3.5%)	3 (#3) (2.3%)	2 (1.3%)	4 (1.6%)
SCU + TAKE + NP + as-NP	0 (0%)	1 (0.9%)	0 (0%)	0 (0%)	7 (#4) (2.7%)
SCU + TAKE + up + NP	0 (0%)	1 (0.9%)	0 (0%)	3 (#4) (1.9%)	5 (#5) (2.0%)
[it] + TAKE + NP + for-NP + to-inf	0 (0%)	0 (0%)	0 (0%)	5 (#3) (3.2%)	0 (0%)
[it] + TAKE + NP + NP + to-inf	0 (0%)	2 (1.8%)	4 (#2) (4.6%)	0 (0%)	0 (0%)
Cumulative % (top 5)	69.5%	90.3%	87.4%	93.6%	86.8%

Table A3. Senses of TAKE in writing in each variety (raw frequencies, rankings, and percentages).

	NativeE	BFE	MCE	HKE	SgE
do (delexical sense)	110 (#1) (33.4%)	123 (#1) (32.8%)	93 (#2) (21.5%)	205 (#1) (40.2%)	86 (#2) (22.8%)
idioms	44 (#2) (13.4%)	60 (#2) (16.0%)	6 (1.4%)	10 (2.0%)	19 (#4) (5.0%)
particle verbs	41 (#3) (12.5%)	34 (#5) (9.1%)	45 (#3) (10.4%)	162 (#2) (31.8%)	197 (#1) (52.1%)
capture	25 (#4) (7.6%)	19 (5.1%)	10 (2.3%)	2 (1.6%)	6 (0.4%)
require	25 (#4) (7.6%)	39 (#4) (10.4%)	16 (3.7%)	14 (#5) (2.7%)	12 (#5) (3.2%)
consider	18 (5.5%)	52 (#3) (13.9%)	32 (#4) (7.4%)	7 (1.4%)	8 (2.1%)
move	17 (5.2%)	8 (2.1%)	3 (0.7%)	13 (2.5%)	7 (1.9%)
assume	11 (3.3%)	14 (3.7%)	25 (#5) (5.8%)	27 (#4) (5.3%)	9 (2.4%)
engage in	10 (3.0%)	3 (0.8%)	10 (2.3%)	2 (0.4%)	3 (0.8%)
accept	9 (2.7%)	2 (0.5%)	164 (#1) (38.0%)	32 (#3) (6.3%)	23 (#3) (6.1%)
habitual actions and qualities	8 (2.4%)	8 (2.1%)	16 (3.7%)	14 (2.7%)	2 (0.5%)
use	3 (0.9%)	3 (0.8%)	0 (0.0%)	6 (1.2%)	4 (1.1%)
travel	0 (0.0%)	2 (0.5%)	1 (0.2%)	5 (1.0%)	0 (0.0%)
grab	0 (0.0%)	3 (0.8%)	1 (0.2%)	0 (0.0%)	0 (0.0%)
other	8 (2.4%)	5 (1.3%)	10 (2.3%)	11 (2.2%)	3 (0.8%)
Cumulative % (top 5)	74.5%	82.2%	83.1%	86.3%	89.2%
Total	329 100%	375 100%	432 100%	510 100%	378 100%

Table A4. Senses of TAKE in speech in each variety (raw frequencies, rankings, and percentages).

	NativeE	BFE	MCE	HKE	SgE
particle verbs	46 (#1) (28.4%)	6 (5.3%)	7 (#5) (8.0%)	8 (#5) (5.1%)	42 (#3) (16.3%)
require	36 (#2) (22.2%)	18 (#2) (15.9%)	15 (#2) (17.2%)	17 (#3) (10.8%)	17 (#4) (6.6%)
move	30 (#3) (18.5%)	8 (#5) (7.1%)	10 (#4) (11.5%)	8 (#5) (5.1%)	9 (3.5%)
do (delexical sense)	14 (#4) (8.6%)	18 (#2) (15.9%)	20 (#1) (23.0%)	40 (#2) (25.5%)	55 (#2) (21.4%)
engage in	9 (#5) (5.6%)	5 (4.4%)	15 (#2) (17.2%)	42 (#1) (26.8%)	66 (#1) (25.7%)
idioms	5 (3.1%)	1 (0.9%)	0 (0.0%)	5 (3.2%)	3 (1.2%)
assume	3 (1.9%)	3 (2.7%)	1 (1.1%)	1 (0.6%)	6 (2.3%)
consider	3 (1.9%)	2 (1.8%)	1 (1.1%)	1 (0.6%)	8 (3.1%)
accept	3 (1.9%)	7 (6.2%)	3 (3.4%)	7 (4.5%)	5 (1.9%)
capture	3 (1.9%)	2 (1.8%)	2 (2.3%)	1 (0.6%)	10 (3.9%)
travel	2 (1.2%)	0 (0.0%)	0 (0.0%)	16 (#4) (10.2%)	21 (#4) (8.2%)
grab	2 (1.2%)	1 (0.9%)	0 (0.0%)	1 (0.6%)	2 (0.8%)
use	1 (0.6%)	29 (#1) (25.7%)	4 (4.6%)	4 (2.5%)	7 (2.7%)
habitual actions and qualities	1 (0.6%)	4 (3.5%)	4 (4.6%)	1 (0.6%)	2 (0.8%)
other	4 (2.5%)	9 (#4) (8.0%)	5 (5.7%)	5 (3.2%)	4 (1.6%)
Cumulative % (top 5)	83.3%	72.6%	76.9%	78.4%	78.2%
Total	162 (100%)	113 (100%)	87 (100%)	157 (100%)	257 (100%)

Notes

- ¹ In the literature, alternative terms like English as a second language (e.g., [Gilquin & Granger, 2011](#)), postcolonial Englishes (e.g., [Schneider, 2007](#)), indigenized varieties of English (e.g., [Sridhar & Sridhar, 1986](#)), to name a few, are used to refer to NEs, while English as a foreign language (e.g., [Gilquin & Granger, 2011](#)), performance varieties (e.g., [Kachru, 1982](#)), and non-postcolonial Englishes (e.g., [Buschfeld et al., 2018](#)) are used to refer to LEs. In this article, LEs and NEs are chosen because the parallel terminology enables systematic comparison while plural forms reflect the diversity of English varieties.
- ² While English mass media and new communication technologies have created unprecedented opportunities for English exposure beyond classroom settings in LEs contexts, there might be a gap between availability and actual use of these resources. For instance, in mainland China, despite the availability of English mass media (e.g., *China Daily*) and English materials (e.g., English films and books), learners of English rarely exploit these resources ([Wei & Su, 2012](#); [Zheng, 2014](#)).
- ³ This includes “Second-Language Varieties and Learner Englishes” at the First Conference of the International Society for the Linguistics of English in 2008 ([Mukherjee & Hundt, 2011](#)), “Corpus Linguistics and Linguistic Innovations in Non-native Englishes” at the ICAME 36 conference in 2015 ([Deshors et al., 2018](#)), and “Global Englishes and SLA: Establishing a Dialogue and Common Research Agenda” at the American Association for Applied Linguistics conference in 2016 ([Bolton & De Costa, 2018](#)).
- ⁴ Despite the high comparability of these three corpora, one caveat to bear in mind is that they were collected at different times (see [Gilquin, 2024](#), p. 5). For instance, LINDSEI-MC was compiled in 2001, while NESSI-HK was collected between 2016 and 2017.
- ⁵ For the full list of complement types, please refer to the VDE ([Herbst et al., 2004](#)).
- ⁶ SCU stands for subject complement unit. In the VDE, subjects are not specified in the valency patterns to maintain descriptive simplicity. For instance, the pattern of TAKE in (3) is described as + TAKE + NP + as-NP in the VDE. However, the subject will be

- represented to provide a more complete account of valency patterns in this study. Given the difficulty of identifying SCU in an unparsed corpus, valency patterns are usually presented in a reduced form (e.g., SCU + TAKE + NP) (see also Faulhaber, 2011).
- 7 Rather than using labels like “require”, the VDE adopts a less systematic and more flexible lexicographical approach. For instance, “someone can take a certain amount of time to do something, i.e., need that time to do it” in VDE corresponds to the formal label “require” in Gilquin (2008).
- 8 It shall be noted that this classification is not purely semantic, as categories such as *particle verbs* and *other* do not represent specific senses of TAKE.
- 9 The ComplexHeatmap package in R was used to generate these plots <https://jokergoo.github.io/ComplexHeatmap-reference/book/index.html> (accessed on 1 March 2025).
- 10 They used Gilquin’s (2008) semantic categories of TAKE, which are comparable to the present study.

References

- Algeo, J. (2006). *British or American English? A handbook of word and grammar patterns*. Cambridge University Press.
- Baker, P. (2017). *American and British English: Divided by a common language?* Cambridge University Press.
- Bernaish, T., & Götz, S. (2021). Across three Kachruvian circles with two parts-of-speech: Nouns and verbs in ENL, ESL, and EFL varieties. In P. Peters, & K. Burridge (Eds.), *Exploring the ecology of world Englishes in the twenty-first century: Language, society and culture* (pp. 215–237). Edinburgh University Press.
- Biber, D. (1991). *Variation across speech and writing*. Cambridge University Press.
- Biewer, C. (2011). Modal auxiliaries in second language varieties of English: A learner’s perspective. In J. Mukherjee, & M. Hundt (Eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap* (pp. 7–33). John Benjamins.
- Bolton, K. (2003). *Chinese Englishes: A sociolinguistic history*. Cambridge University Press.
- Bolton, K., & De Costa, P. I. (2018). World Englishes and second language acquisition: Introduction. *World Englishes*, 37(1), 2–4. [CrossRef]
- Brezina, V., Weill-Tessier, P., & McEnery, A. (2021). *LancsBox v. 6. x. [software package]*. Lancaster University.
- Buschfeld, S. (2013). *English in Cyprus or Cyprus English: An empirical investigation of variety status*. John Benjamins.
- Buschfeld, S., & Kautzsch, A. (2014). English in Namibia. *English World-Wide*, 35(2), 121–160. [CrossRef]
- Buschfeld, S., & Kautzsch, A. (2017). Towards an integrated approach to postcolonial and non-postcolonial Englishes. *World Englishes*, 36(1), 104–126. [CrossRef]
- Buschfeld, S., & Kautzsch, A. (Eds.). (2020). *Modelling world Englishes: A joint approach to postcolonial and non-postcolonial varieties*. Edinburgh University Press.
- Buschfeld, S., Kautzsch, A., & Schneider, E. (2018). From colonial dynamism to current transnationalism: A unified view on post-colonial and non-postcolonial Englishes. In S. Deshors (Ed.), *Modeling world Englishes: Assessing the interplay of emancipation and globalization of ESL varieties* (pp. 15–44). John Benjamins.
- Callies, M. (2016). Towards a process-oriented approach to comparing EFL and ESL varieties: A corpus study of lexical innovations. *International Journal of Learner Corpus Research*, 2(2), 229–250. [CrossRef]
- Croft, W. (2000). *Explaining language change: An evolutionary approach*. Longman.
- Davydova, J. (2012). Englishes in the outer and expanding circles: A comparative study. *World Englishes*, 31(3), 366–385. [CrossRef]
- Davydova, J. (2019). *Quotation in indigenised and learner English: A sociolinguistic account of variation*. De Gruyter Mouton.
- De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures (BELL), New Series*, 2, 225–246.
- Deshors, S. C. (2014). A case for a unified treatment of EFL and ESL. *English World-Wide. A Journal of Varieties of English*, 35(3), 277–305. [CrossRef]
- Deshors, S. C., Götz, S., & Laporte, S. (Eds.). (2018). *Rethinking linguistic creativity in non-native Englishes*. John Benjamins.
- Deshors, S. C., & Gries, S. T. (2016). Profiling verb complementation constructions across new Englishes: A two-step random forests analysis of ing vs. to complements. *International Journal of Corpus Linguistics*, 21(2), 192–218. [CrossRef]
- Diessel, H. (2014). Usage-based linguistics. In M. Aronoff (Ed.), *Oxford bibliographies in linguistics*. Oxford University Press.
- Dixon, R. M. W. (2005). *A semantic approach to English grammar*. Oxford University Press.
- Edwards, A. (2014). *English in the Netherlands: Functions, forms and attitudes* [Unpublished doctoral thesis, University of Cambridge].
- Edwards, A., & Lange, R. J. (2016). In case of innovation: Academic phraseology in the three circles. *International Journal of Learner Corpus Research*, 2(2), 252–277. [CrossRef]
- Edwards, A., & Laporte, S. (2015). Outer and expanding circle Englishes: The competing roles of norm orientation and proficiency levels. *English World-Wide*, 36(2), 135–169. [CrossRef]
- Eskildsen, S. W. (2009). Constructing another language—Usage-based linguistics in second language acquisition. *Applied Linguistics*, 30(3), 335–357. [CrossRef]
- Faulhaber, S. (2011). *Verb valency patterns: A challenge for semantics-based accounts*. De Gruyter Mouton.

- Gilquin, G. (2008). What you think ain't what you get: Highly polysemous verbs in mind and language. In J.-R. Lapaire, G. Desagulier, & J.-B. Guignard (Eds.), *Du fait grammatical au fait cognitif. From gram to mind: Grammar as cognition* (Vol. 2, pp. 235–255). Presses Universitaires de Bordeaux.
- Gilquin, G. (2011). Corpus linguistics to bridge the gap between world Englishes and learner Englishes. In L. R. Miyares, & M. R. Á. Silva (Eds.), *Comunicación Social en el siglo XXI* (Vol. II, pp. 638–642). Centro de Lingüística Aplicada.
- Gilquin, G. (2015a). At the interface of contact linguistics and second language acquisition research: New Englishes and Learner Englishes compared. *English World-Wide*, 36(1), 91–124. [\[CrossRef\]](#)
- Gilquin, G. (2015b). The use of phrasal verbs by French-speaking EFL learners. A constructional and collostructional corpus-based approach. *Corpus Linguistics and Linguistic Theory*, 11(1), 51–88. [\[CrossRef\]](#)
- Gilquin, G. (2016a). Input-dependent L2 acquisition: Causative constructions in English as a foreign and second language. In S. De Knop, & G. Gilquin (Eds.), *Applied construction grammar* (pp. 115–148). De Gruyter Mouton.
- Gilquin, G. (2016b). Discourse markers in L2 English: From classroom to naturalistic input. In O. Timofeeva, A.-C. Gardner, A. Honkapohja, & S. Chevalier (Eds.), *New approaches to English linguistics: Building bridges* (pp. 213–249). John Benjamins.
- Gilquin, G. (2024). Lexical use in spoken new Englishes and learner Englishes: The effects of shared and distinct communicative constraints. In B. Van Rooy, & H. Kotze (Eds.), *Constraints on language variation and change in complex multilingual contact settings* (pp. 120–152). John Benjamins.
- Gilquin, G. (2025). Second and foreign language learners: The effect of language exposure on the use of English phrasal verbs. *International Journal of Bilingualism*, 29(2), 456–473. [\[CrossRef\]](#)
- Gilquin, G., De Cock, S., & Granger, S. (2010). *Louvain international database of spoken English interlanguage. Handbook and CD-ROM*. Presses Universitaires de Louvain.
- Gilquin, G., & Granger, S. (2011). From EFL to ESL: Evidence from the international corpus of learner English. In J. Mukherjee, & M. Hundt (Eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap* (pp. 55–78). John Benjamins.
- Gilquin, G., & Granger, S. (2021). The passive and the lexis-grammar interface: An inter-varietal perspective. In S. Granger (Ed.), *Perspectives on the L2 phrasicon: The view from learner corpora* (pp. 72–98). Multilingual Matters.
- Gilquin, G., & Meriläinen, L. (2024). Constrained communication in EFL and ESL: The case of embedded inversion. *English World-Wide*, 45(2), 196–223. [\[CrossRef\]](#)
- Giparaitė, J. (2016). Complementation of light verb constructions in world Englishness: A corpus-based study. *Žmogus ir Žodis*, 18(3), 19–39. [\[CrossRef\]](#)
- Görlach, M. (2002). *Still more Englishes*. John Benjamins.
- Götz, S. (2015, August 8–9). *Fluency in ENL, ESL and EFL: A corpus-based pilot study*. Proceedings of Disfluency in Spontaneous Speech, DISS 2015, Edinburgh, UK.
- Götz, S., & Schilk, M. (2011). Formulaic sequences in spoken ENL, ESL and EFL: Focus on British English, Indian English and learner English of advanced German learners. In J. Mukherjee, & M. Hundt (Eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap* (pp. 79–100). John Benjamins.
- Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 3–18). Addison Wesley Longman.
- Granger, S., Dupont, M., Meunier, F., Naets, H., & Paquot, M. (2020). *The international corpus of learner English (Version 3)*. Presses Universitaires de Louvain.
- Greenbaum, S. (1988). A proposal for an international computerized corpus of English. *World Englishes*, 7(3), 315. [\[CrossRef\]](#)
- Gries, S. T., & Deshors, S. C. (2015). EFL and/or ESL? A multi-level regression modeling perspective on bridging the paradigm gap. *International Journal of Learner Corpus Research*, 1(1), 130–159. [\[CrossRef\]](#)
- Hanks, P. (2013). *Lexical analysis: Norms and exploitations*. MIT Press.
- Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4(2), 237–258. [\[CrossRef\]](#)
- He, D., & Li, D. C. S. (2009). Language attitudes and linguistic features in the “China English” debate. *World Englishes*, 28(1), 70–89. [\[CrossRef\]](#)
- Herbst, T., Heath, D., Roe, I., & Götz, D. (2004). *A valency dictionary of English: A corpus-based analysis of the complementation patterns of English verbs, nouns and adjectives*. De Gruyter Mouton.
- Hilbert, M. (2011). Interrogative inversion as a learner phenomenon in English contact varieties: A case of Angloverbals? In J. Mukherjee, & M. Hundt (Eds.), *Exploring second-language varieties of English and Learner Englishes: Bridging a paradigm gap* (pp. 125–143). John Benjamins.
- Ishikawa, S. (2023). *The ICNALE guide: An introduction to a learner corpus study on Asian learners' L2 English*. Routledge.
- Kachru, B. B. (Ed.). (1982). Models for Non-Native Englishes. In *The Other Tongue: English across Cultures* (pp. 31–57). Pergamon Press.
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk, & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge University Press.

- Laporte, S. (2012). Mind the gap! Bridge between world Englishes and learner Englishes in the making. *English Text Construction*, 5(2), 264–291. [\[CrossRef\]](#)
- Laporte, S. (2021). *Corpora, constructions, new Englishes: A constructional and variationist approach to verb patterning*. John Benjamins.
- Leech, G. (2000). Grammars of spoken English: New outcomes of corpus-oriented research. *Language Learning*, 50(4), 675–724. [\[CrossRef\]](#)
- Levshina, N. (2015). *How to do linguistics with R. Data exploration and statistical analysis*. John Benjamins.
- Liu, E. T. K., & Shaw, P. M. (2001). Investigating learner vocabulary: A possible approach to looking at EFL/ESL learners' qualitative knowledge of the word. *IRAL—International Review of Applied Linguistics in Language Teaching*, 39(3), 171–194. [\[CrossRef\]](#)
- Lowenberg, P. H. (1986). Non-native varieties of English: Nativization, norms, and implications. *Studies in Second Language Acquisition*, 8(1), 1–18. [\[CrossRef\]](#)
- Ma, Q., & Xu, Z. (2017). The nativization of English in China. In Z. Xu, D. He, & D. Deterding (Eds.), *Researching Chinese English: The state of the art* (pp. 189–201). Springer.
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. [\[CrossRef\]](#)
- Meriläinen, L. (2017). The progressive form in learner Englishes: Examining variation across corpora. *World Englishes*, 36(4), 760–783. [\[CrossRef\]](#)
- Meunier, F. (2020). Status of English in Belgium. In S. Granger, M. Dupont, F. Meunier, H. Naets, & M. Paquot (Eds.), *International corpus of learner English version 3* (pp. 180–185). Presses Universitaires de Louvain.
- Mežek, Š. (2024). English in Sweden: Functions, features and debates. *World Englishes*, 43(2), 332–345. [\[CrossRef\]](#)
- Miller, J., & Weinert, R. (1998). *Spontaneous spoken language: Syntax and discourse*. Oxford University Press.
- Mukherjee, J., & Gries, S. T. (2009). Collostructional nativisation in new Englishes: Verb-construction associations in the international corpus of English. *English World-Wide*, 30(1), 27–51. [\[CrossRef\]](#)
- Mukherjee, J., & Hoffmann, S. (2006). Describing verb-complementational profiles of new Englishes: A pilot study of IndE. *English World-Wide*, 27(2), 147–173. [\[CrossRef\]](#)
- Mukherjee, J., & Hundt, M. (Eds.). (2011). *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap*. John Benjamins.
- Nesselhauf, N. (2009). Co-selection phenomena across new Englishes: Parallels (and differences) to foreign learner varieties. *English World-Wide*, 30(1), 1–25. [\[CrossRef\]](#)
- Paulasto, H., & Meriläinen, L. (2023). The processes of preposition omission across English variety types. In P. Rautioaho, H. Parviainen, M. Kaunisto, & A. Nurmi (Eds.), *Social and regional variation in world Englishes: Local and global perspectives* (pp. 91–122). Routledge.
- Percillier, M. (2016). *World Englishes and second Language acquisition: Insights from southeast Asian Englishes*. John Benjamins.
- R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Romasanta, R. P. (2020). *Variation in the clausal complementation system in world Englishes: A corpus-based study of regret* [Unpublished doctoral dissertation, Universidade de Vigo].
- Schneider, E. W. (2004). How to trace structural nativization: Particle verbs in world Englishes. *World Englishes*, 23(2), 227–249. [\[CrossRef\]](#)
- Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. Cambridge University Press.
- Schneider, E. W. (2012). Exploring the interface between world Englishes and second language acquisition—And implications for English as a lingua franca. *Journal of English as a Lingua Franca*, 1(1), 57–91. [\[CrossRef\]](#)
- Schneider, E. W. (2014). New reflections on the evolutionary dynamics of world Englishes. *World Englishes*, 33(1), 9–32. [\[CrossRef\]](#)
- Sridhar, K. K., & Sridhar, S. N. (1986). Bridging the paradigm gap: Second language acquisition theory and indigenized varieties of English. *World Englishes*, 5(1), 3–14. [\[CrossRef\]](#)
- Szmrecsanyi, B., & Kortmann, B. (2011). Typological profiling: Learner Englishes versus indigenized L2 varieties of English. In J. Mukherjee, & M. Hundt (Eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap* (pp. 167–207). John Benjamins.
- Van Rooy, B. (2011). A principled distinction between error and conventionalized innovation in African Englishes. In J. Mukherjee, & M. Hundt (Eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap* (pp. 189–208). John Benjamins.
- Wang, Y. (2016). *The idiom principle and L1 influence: A contrastive learner-corpus study of delexical verb + noun collocations*. John Benjamins.
- Wei, R., & Su, J. (2012). The statistics of English in China. *English Today*, 28(3), 10–14. [\[CrossRef\]](#)
- Werner, J., & Mukherjee, J. (2012). Highly polysemous verbs in new Englishes: A corpus-based study of Sri Lankan and Indian English. In S. Hoffmann, P. Rayson, & G. Leech (Eds.), *Corpus linguistics: Looking back—Moving forward* (pp. 249–266). Rodopi.
- Williams, J. (1987). Non-native varieties of English: A special case of language acquisition. *English World-Wide*, 8(2), 161–199. [\[CrossRef\]](#)
- Xu, Z. (2010). *Chinese English: Features and implications*. Open University of Hong Kong Press.
- Zhao, Q., & Jiang, J. (2020). Verb valency in interlanguage: An extension to valency theory and new perspective on L2 learning. *Poznań Studies in Contemporary Linguistics*, 56(2), 339–363. [\[CrossRef\]](#)

- Zheng, Y. (2014). A phantom to kill: The challenges for Chinese learners to use English as a global language. *English Today*, 30(4), 34–39. [[CrossRef](#)]
- Zipp, L., & Bernaisch, T. (2012). Particle verbs across first and second language varieties of English. In M. Hundt, & U. Gut (Eds.), *Mapping unity and diversity worldwide: Corpus-based studies of new Englishes* (pp. 167–196). John Benjamins.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.