

It's Look, That's Go: Usage-Based Innovation in Asian Englishes

Zayani Zainal Abidin
Universiti Brunei Darussalam
zayaniabidin93@gmail.com

Article History:

First Received:
26/11/2025

Final Revision:
20/06/2026

Available online:
30/06/2026

Abstract. This study extends previous research on the *it's + Vø* pattern in Asian English as a lingua franca (ELF) by presenting the first comparative analysis of the *it's + Vø* (e.g., *it's look* for 'it looks') and *that's + Vø* constructions (e.g., *that's go* for 'that goes'). It examines how usage frequency, functional motivations, and cognitive mechanisms shape constructional innovation across three English domains: institutionalised varieties, ELF interactions, and learner Englishes. Using a mixed-method approach that combines frequency and distributional analysis, collocational and collostructional profiling, and qualitative discourse analysis, the study systematically compares the constructions to identify structural and contextual motivations underlying their use. Findings show that *it's + Vø* is more frequent, lexically diverse, and socially widespread than *that's + Vø*, with highest productivity in learner Englishes, moderate entrenchment in ELF use, and marginal presence in institutionalised varieties. Analysis of recurring bare-verb patterns and functional clustering indicates fluid and context-responsive usage in ELF interactions, whereas learners' preference for evaluative or stance-marking expressions reflects pragmatic overextension. These patterns suggest ongoing processes of diffusion, schematic extension, and functional adaptation rather than stable conventionalisation. The study contributes to usage-based accounts of grammatical innovation in Asian Englishes and reframes these constructions as functionally motivated rather than erroneous.

Keywords: *emergent constructions; English as a Lingua Franca; learner English; Asian English morphosyntax; Construction Grammar*

INTRODUCTION

In canonical English, contracted forms such as *it's* and *that's* function as subject-copula units, where *s* represents the verb *is* and selects complements appropriate to the copular construction (Biber et al., 1999, pp. 1005–6). However, several studies working from different methodological approaches and theoretical orientations have reported instances in which *it's* or *that's* does not behave as a decomposable 'subject + copula' sequence. Instead, these forms appear as holistic, prefabricated subject units that can combine directly with a bare or uninflected verb (*V*Ø). For example, in a one-time task-based study of Dutch learners of English, Poulisse (1999, pp. 227–230) reported the use of *it's* + *V*Ø in *it's belong* ('it belongs') surfacing in her findings and interpreted such usage as slips of the tongue. In a longitudinal study on the morphosyntactic performance of an advanced Malaysian L2-English speaker, Mohd Yusoff et al. (2019, pp. 165–166) uncovered a consistent use of *it's look* ('it looks'), *it's mean* ('it means'), and *that's* + *V*Ø in *that's mean* ('that means'), and classified such usage as sporadic errors. A similar stance is evident in Hamid and Hasan's (2020) discussion of Bangladeshi English features, where the authors directly cited the use of "*that's*, " *meaning* " *that means* ", in a Bangladeshi magazine editorial as linguistic carelessness:

We may did something wrong during this year or still making mistakes, *that's mean* [emphasis added] we are making new things, trying new things, learning, living, pushing ourselves, changing ourselves, changing our world. (p. 311)

Although these error-based interpretations provide insight into L2-English morphosyntactic performance, they raise a question of whether these forms may instead index emerging, systematic speech patterns.

Recent corpus-based evidence supports this possibility, suggesting that these constructions are not random or incidental errors but may reflect underlying cognitive processes. Zayani (2024) is the first study to specifically investigate the use of the *it's* + *V*Ø sequence in ELF interactions from the Asian Corpus of English (ACE), proposing that priming and chunking of canonical *it's* forms, combined with the variable English inflection, facilitate its emergence. While this work advances the discussion beyond prescriptive error-based interpretations, its empirical and theoretical scope remains limited. In particular, it does not incorporate other relevant cognitive mechanisms such as analogy, categorisation, and frequency effects (Bybee, 2010), nor does it consider processing pressures such as retrieval interference (Van Dyke & McElree, 2006). The study also overlooks experience-based social factors, including the role of speaker demographics, interactional roles, and communicative contexts in shaping constructional choice (Díaz-Campos & Pollock, 2023). Without integrating these cognitive and sociolinguistic dimensions, the analysis provides only a partial account of the processes driving the emergence and diffusion of *it's* + *V*Ø in ELF settings.

Although these previous studies have documented instances of *it's* + *V*Ø and *that's* + *V*Ø only in a small number of L2-English contexts (Dutch, Malaysian, and Bangladeshi) and in Asian ELF interactions, it remains unclear whether these forms should be understood as merely sporadic learner errors or as emerging, systematic constructions. This uncertainty is compounded by the lack of research examining the cognitive and functional mechanisms that may underpin their use, including the roles of analogy, frequency-based entrenchment, and the possible reanalysis of *it's* and *that's* as holistic stance-prefacing units. Furthermore, the

ways in which different usage contexts, such as institutionalised varieties, ELF interactions, and learner Englishes, shape the frequency, lexical range, and interactional functions of these constructions have not been systematically investigated. These issues indicate that the phenomenon is not yet well understood either structurally or functionally, nor in terms of the contextual pressures that may promote its emergence.

Despite the empirical and theoretical relevance of the *it's + Vø* and *that's + Vø* constructions, three major gaps remain. First, no study has directly compared ELF, institutionalised and learner Englishes corpora to determine whether these constructions exhibit consistent patterns across contexts or reflect environment-specific developmental trajectories. Second, their collocational and collostructional behaviour has not been systematically analysed, leaving unclear whether particular verbs or semantic domains (e.g., evaluation, stance, procedural meanings) drive their usage. Third, the functional motivations underlying these constructions, whether shaped by stance marking, discourse alignment, or frequency-based entrenchment, have yet to be examined. Addressing these gaps requires a comparative, multi-corpus approach that can reveal how the constructions behave across contexts and illuminate the structural, functional, and cognitive processes driving their emergence.

To address these empirical and theoretical gaps, the present study undertakes a comparative, multi-corpus investigation of the emergent *it's + Vø* and *that's + Vø* constructions, focusing on Asian English contexts to extend previous documentation, which has so far reported these constructions primarily in Malaysian, Bangladeshi, and Asian ELF settings. By examining how these forms pattern across institutionalised varieties (International Corpus of English; ICE), ELF interactions (ACE), and learner Englishes (International Corpus Network of Asian Learners of English; ICNALE), the study seeks to establish whether they function as systematic constructions rather than sporadic deviations. This comparative perspective provides an integrated basis for explaining how high-frequency subject-copula units may be reanalysed as productive templates and how structural, functional, and cognitive pressures shape their emergence.

Accordingly, this study is guided by four objectives:

1. To provide a comparative account of the emergent *it's + Vø* and *that's + Vø* constructions across institutionalised varieties, ELF interactions, and learner Englishes.
2. To examine the structural, functional, and cognitive motivations underlying these constructions, including frequency patterns, lexical distributions, and collocational and semantic profiles.
3. To investigate and compare how different usage contexts shape constructional behaviour.
4. To contribute to usage-based and constructional accounts of grammatical innovation in Asian Englishes by demonstrating how high-frequency subject-copula units (*it's*, *that's*) can serve as templates for emergent syntactic patterns.

Beyond these objectives, the study's novelty lies in its integrated analytical framework and expanded empirical scope. It provides the first cross-domain comparative analysis of both the *it's + Vø* and *that's + Vø* constructions in Asian Englishes, drawing together insights from Construction Grammar and usage-based sociolinguistics. By documenting the previously unexamined *that's + Vø* pattern and extending the analysis to Asian institutionalised varieties and Asian learner Englishes, the study develops a comprehensive

account of the structural, functional, and interactional dimensions of these emergent constructions. Its combination of quantitative corpus analysis with qualitative discourse examination links micro-level usage patterns to broader processes of variation and change, yielding new insights into how high-frequency subject–copula units become productive templates in Asian Englishes.

Theoretically, the study advances usage-based and constructional accounts of grammatical innovation by providing the first comparative evidence for how *it's* + *V*Ø and *that's* + *V*Ø behave across institutionalised, ELF, and learner Englishes. It identifies *that's* + *V*Ø as an emergent pattern, demonstrates how frequent subject–copula units may be reanalysed as holistic templates that license schematic extension, and supplies new empirical support for the roles of entrenchment, analogy, and processing economy in the development of non-canonical constructions. The contrasting patterns across the three domains refine our understanding of how context-sensitive interactional pressures shape constructional innovation in Asian Englishes.

In practice, the study reframes these forms as functionally motivated, context-dependent constructions rather than categorical errors, with implications for ELF-informed pedagogy, learner assessment, and the recognition of grammatical diversity in English. Methodologically, its triangulated mixed-method design provides a replicable model for future research on emergent constructions and usage-based grammatical change.

To examine the emergence, distribution, and stability of the *it's* + *V*Ø and *that's* + *V*Ø constructions, this study integrates two complementary perspectives: (i) Construction Grammar and usage-based cognitive linguistics, which explain how constructions arise and become entrenched; and (ii) usage-based sociolinguistics, which accounts for their selective diffusion across communities.

Constructional Emergence in a Usage-Based Framework

Construction Grammar (CxG) provides the primary theoretical lens for analysing the *it's* + *V*Ø and *that's* + *V*Ø sequences, conceptualising linguistic knowledge as a network of form–meaning pairings that range from highly specific lexical units to fully schematic patterns (Goldberg, 1995, 2006). Within this framework, constructions emerge from patterns in usage. This emergence is gradual, experience-driven, and grounded in domain-general cognitive processes such as categorisation, chunking, and analogy (Bybee, 2010).

Reduced forms such as *it's* and *that's* are particularly susceptible to chunking (see Vetchinnikova & Hiltunen, 2021, on *it's*), a process well documented for other English contractions (e.g., Bybee & Scheibman, 1999, on *don't*; Levshina & Lorenz, 2022, on *wanna*). Chunking encourages speakers to treat these strings as single processing units (Ellis, 2002, 2017). As *it's* and *that's* occur frequently in canonical expressions such as *it's not too bad* or *that's a good idea* (Biber et al., 1999, pp. 1005–6), they may become entrenched not merely as decomposable 'subject + copula' units, but as prefabricated discourse-prefacing constructions functioning as holistic cues for stance, evaluation, or frame-setting. As a result, schematic templates (e.g., '*it's* + predicate' and '*that's* + predicate') become well established in speakers' mental grammars.

When analogy enables speakers to map familiar patterns onto new verbs and discourse functions (Bybee, 2010), these entrenched templates can occasionally extend into non-canonical environments, such as *it's look* and *that's go*, paralleling canonical meanings *it looks*

and *that goes*. These sequences are consistent with accounts of constructional reanalysis in which high-frequency entrenched templates undergo functional extension and develop into productive, novel syntactic patterning acquiring new pragmatic and semantic values (Bybee, 2010).

Because usage-based theories emphasise experience-driven learning, the surrounding linguistic ecology (whether ELF interactions, institutional norms, or learner input) ultimately shapes how such emerging constructions arise, stabilise, or vary across contexts. This framework provides the basis for analysing the emergent *it's + Vø* and *that's + Vø* patterns across the corpora examined in the present study.

Sociolinguistic Conditioning in a Usage-Based Variation

While cognitive processes explain *how* constructions emerge, a usage-based sociolinguistic perspective helps account for *why* particular constructions are selected, extended, or restricted across speech communities. Constructional choices are reflected not only by frequency and processing factors but also by speaker demographics, interactional roles, participation frameworks, and community norms (Díaz-Campos & Pollock, 2023). From this viewpoint, emergent patterns reflect both experience-driven learning under specific social and communicative conditions.

In ELF contexts, interaction is guided by goals such as intelligibility, efficiency, and alignment (Mauranen, 2012, 2018). These communicative pressures favour reduced, prefabricated, and hybridised constructions that facilitate real-time negotiation of meaning. Because adherence to 'standard' morphosyntax is often unnecessary for successful communication (Seidlhofer, 2004), features such as absence of the third-person singular present tense -s (e.g., *he say*), redundant prepositions (e.g., *study about*), or semantically transparent repetitions (e.g., *black colour*) may occur without impairing understanding. Within this ecology, familiar templates such as '*it's + predicate*' or '*that's + predicate*' may plausibly extend to non-canonical realisations of the *it's + Vø* and *that's + Vø* patterns, as speakers prioritise ease of processing and shared understanding over standardised forms.

Institutionalised second-language varieties present a different usage profile. In these settings, widespread intracommunity exposure, community-level entrenchment, and long-term contact may stabilise particular constructional patterns (Mufwene, 2010; Schneider, 2007). At the same time, codification, educational standards, and normative expectations may channel or constrain innovation (Schneider, 2007). Such varieties may therefore exhibit selective retention or routinisation of emerging constructions, reflecting both entrenched usage and normative pressures.

Learner contexts represent yet another usage ecology. Learners often rely heavily on familiar, high-frequency items, which increases the likelihood of overextension or analogical mapping beyond canonical patterns (Ellis, 2002). Restricted input, pedagogical focus on formulaic sequences, and limited exposure to diverse registers shape constructional availability (Ellis, 2002). Consequently, learner Englishes may display innovations arising from systematic, experience-driven learning rather than community-level conventionalisation (Ellis, 1997).

In sum, usage-based sociolinguistics situates constructional emergence within socially and interactionally grounded distributions of experience. This framework predicts that systematic differences in the frequency, functions, and meanings of *it's + Vø* and *that's + Vø*

will vary systematically across ELF interactions, institutionalised varieties and learner Englishes.

RESEARCH METHOD

Corpora

International Corpus of English (ICE)

ICE is a one-million-word corpus representing spoken and written varieties of institutionalised World Englishes (Greenbaum, 1996). Each national component comprises 300 spoken and 200 written texts of roughly 2,000 words, produced by adult speakers educated in the target variety. Only the spoken components (600,000 words), which are available through the official ICE website (ICE, n.d.), are used in this study. These cover a wide range of discourse situations (see Table A in Appendix A).

Asian Corpus of English (ACE)

ACE (Wang & Kirkpatrick, 2024) is a one-million-word corpus of naturally occurring spoken ELF interactions recorded across nine Asian locations between 2009 and 2014. The data span academic, workplace, and informal settings and include rich speaker metadata (age, sex, L1, nationality, education, occupation). The ACE online interface provides access to 145 files categorised by country, domain, and topic (ACE, 2024).

International Corpus Network of Asian Learners of English (ICNALE)

ICNALE is a multi-country learner corpus containing over 15,000 spoken and written productions from more than 5,500 learners (plus native-speaker controls) across East, Southeast, and South Asia (Ishikawa, 2023). Only the 2.1-million-word spoken datasets are used here, consisting of 60-second monologues and 30–40-minute semi-structured interviews on two everyday ICNALE topics ('part-time jobs for college students' and 'non-smoking in restaurants'). Although elicited in institutional settings, these tasks yield conversational, informal registers, ensuring comparability with informal speech in ICE and ACE. The study uses Spoken Monologue v2.0 and Spoken Dialogue v1.3, the versions available at the time of data extraction (i.e. 2023). These datasets are accessible through the ICNALE online platform maintained by the project team (Ishikawa, 2023).

Including ICE, ACE, and ICNALE provides a balanced dataset spanning established, contact-driven, and learner varieties of English. ICE offers a norm-governed baseline for assessing whether *it's* + *V*Ø and *that's* + *V*Ø reflect broader diffusion beyond institutionalised Englishes. ACE captures ELF interactions where multilingual speakers negotiate meaning dynamically, making it a key site for observing whether these constructions function as adaptive resources in real-time communication. ICNALE complements these by revealing how the constructions surface in learner interlanguage, allowing insights into acquisition, transfer, and emerging conventionalisation.

Data Extraction and Preparation

Data extraction was conducted separately for each corpus. All corpus files were imported into LancsBox (Brezina et al., 2021), where concordance searches were performed for the forms *it's* and *that's*. The resulting concordance lines were manually examined to identify instances of the *it's* + *V*Ø and *that's* + *V*Ø constructions. The search procedure also included

tokens containing intervening elements (e.g., *it's not depend*), reflecting the schematic nature of the constructions as productive frames rather than fixed lexical sequences. For each identified token, the occurrence was extracted together with sufficient surrounding discourse context, and the corresponding speaker and source file were recorded. To facilitate subsequent qualitative analysis, the complete speech segments produced by each selected speaker were then compiled and retained for contextual examination.

ICE Dataset

Tokens occurred only in the Hong Kong, India, and Sri Lanka components, forming a 1.8-million-word ICE dataset. Speaker and file details with newly assigned speaker codes are listed in Table 1. Metadata completeness varies across components (particularly for Sri Lanka), limiting fine-grained sociolinguistic comparison but not structural analysis. Recording periods of the ICE dataset span from the 1990s to 2009 (Hong Kong and India recordings date from the 1990s (Nelson, 2002, 2006); Sri Lanka from 2003 to 2009 (Bernaisch et al., 2019).

Table 1. ICE Dataset (N = 11)

Filenames	Speaker Code	Sex	Age	Nationality	L1	Education	No. of words
S1A-058	ICE-HK-TagF1	F	31–35	Philippines	Tagalog	Secondary	2527
S2A-047	ICE-HK-CanM1	M	36–40	China	Cantonese	University	2282
S2A-053	ICE-HK-CanM2	M	21–25	China	Cantonese	University	2328
S1A-010	ICE-IN-BenF1	F	26–33	India	Bengali	Master's	1115
S1A-016	ICE-IN-TamF2	F	26–33	India	Tamil	Doctorate	789
S1A-001	ICE-IN-MarF3	F	50+	India	Marathi	Master's	777
S1B-026	ICE-LK-XXM1	M	N/A	N/A	N/A	N/A	334
S2A-036	ICE-LK-XXX1	N/A	N/A	N/A	N/A	N/A	2060
S2A-039	ICE-LK-XXM2	M	N/A	N/A	N/A	N/A	2317
S2A-050	ICE-LK-XXM3	M	N/A	N/A	N/A	N/A	2120
S2B-024	ICE-LK-XXM4	M	N/A	N/A	N/A	N/A	2517

Note. Original filenames are included for full reproducibility. Speaker codes follow the format ICE-[component]-[L1, sex, sequence number], where region codes for components are HK = Hong Kong, IN = India, and LK = Sri Lanka; L1 codes are abbreviated forms of the speakers' L1 (e.g., Tag = Tagalog, XX = unknown); sex codes are F = female, M = male, and X = unknown. Sequence numbers are assigned within each nationality–sex combination in ascending order. Due to incomplete metadata in Sri Lanka, the analysis of this component focuses exclusively on structural patterns rather than socio-demographic variation.

ACE Dataset

Relevant tokens were found in the Brunei, China, Hong Kong, Malaysia, the Philippines, Singapore, and Vietnam subsets. Table 2 lists the selected speakers with assigned codes. For full reproducibility, the original filenames of the selected ACE dataset are listed in Table B in Appendix B.

Table 2. ACE Dataset (N = 24)

Speaker Code	Sex	Age	L1	Education	No. of words
ACE-BN-MalM1	M	20–30	Bruneian Malay	Postgraduate	1,627
ACE-CN-ChiF1	F	40–50	Chinese ^a	Postgraduate	905

Table 2 (continued)

Speaker Code	Sex	Age	L1	Education	No. of words
ACE-HK-CanM1	M	40–50	Cantonese	Postgraduate	5,044
ACE-MY-ThaF1	F	20–30	Thai	Graduate	197
ACE-MY-ThaF2	F	20–30	Thai	Postgraduate	441
ACE-MY-MalM1	M	20–30	Malaysian Malay	Postgraduate	984
ACE-MY-MalM2	M	20–30	Malaysian Malay	Graduate	1,541
ACE-MY-ManF1	F	20–30	Mandarin	Postgraduate	3,331
ACE-MY-BurF1	F	40–50	Burmese	Postgraduate	1994
ACE-MY-BurF2	F	30–40	Burmese	Postgraduate	328
ACE-MY-ThaF3	F	20–30	Thai	Postgraduate	1,745
ACE-MY-VieF1	F	30–40	Vietnamese	Postgraduate	715
ACE-MY-KhmM1	M	30–40	Khmer	Graduate	467
ACE-MY-VieF2	F	30–40	Vietnamese	Postgraduate	741
ACE-MY-ThaM1	M	30–40	Thai	Graduate	268
ACE-PH-VieF1	F	20–30	Vietnamese	Graduate	1,101
ACE-PH-BurF1	F	20–30	Burmese	Postgraduate	774
ACE-SG-MalF1	F	20–30	Malaysian Malay	Postgraduate	3,088
ACE-SG-ManF1	F	20–30	Mandarin	Graduate	3,181
ACE-VN-VieF1	F	30–40	Vietnamese	Postgraduate	282
ACE-VN-VieF2	F	30–40	Vietnamese	Postgraduate	1,532
ACE-VN-VieF3	F	20–30	Vietnamese	Postgraduate	2,081
ACE-VN-VieF4	F	20–30	Vietnamese	Postgraduate	1,728
ACE-VN-VieF5	F	<20	Vietnamese	Undergraduate	1,481

Note. Speaker codes follow the format ACE-[recording site]-[L1, sex, sequence number], where recording site codes are BN = Brunei, CN = China, HK = Hong Kong, MY = Malaysia, PH = Philippines, SG = Singapore, and VN = Vietnam; L1 codes are abbreviated forms of the speakers' L1 (e.g., Mal = Malay); sex codes are F = female and M = male; and sequence numbers are assigned within each L1–sex combination in ascending order.

L1 labels follow corpus metadata; *Chinese* is unspecific, whereas varieties (e.g., Cantonese, Mandarin) are indicated where available.

ICNALE Dataset

The ICNALE dataset consists of 61 speakers, drawn from 31 monologue files (29 speakers) and 48 dialogue files (32 speakers). Table 3 provides speaker information with assigned codes. For full reproducibility, the original filenames of the selected ICNALE dataset are listed in Table C in Appendix C.

Table 3. ICNALE Dataset (N = 61)

Speaker Code	Sex	Age	Nationality	Education	Proficiency	L1	No. of words
ICN-Man-M1	M	21	China	Undergraduate	B1_2	Mandarin	401
ICN-Man-M2	M	28	China	Master's	B1_2	Mandarin	427
ICN-Man-F1	F	21	China	Undergraduate	B1_2	Mandarin	503
ICN-Man-F2	F	20	China	Undergraduate	A2_0	Mandarin	432
ICN-Ind-M1	M	21	Indonesia	Undergraduate	B1_2	Indonesian	382
ICN-Ind-M2	M	22	Indonesia	Undergraduate	B1_1	Indonesian	464
ICN-Ind-F1	F	19	Indonesia	Undergraduate	A2_0	Indonesian	330
ICN-Ind-F2	F	25	Indonesia	Undergraduate	B1_1	Indonesian	373

Table 3 (continued)

Speaker Code	Sex	Age	Nationality	Education	Proficiency	L1	No. of words
ICN-Ind-M3	M	21	Indonesia	Undergraduate	B1_1	Indonesian	277
ICN-Ind-F3	F	19	Indonesia	Undergraduate	A2_0	Indonesian	596
ICN-Ind-F4	F	20	Indonesia	Undergraduate	A2_0	Indonesian	501
ICN-Ind-F5	F	20	Indonesia	Undergraduate	B1_2	Indonesian	560
ICN-Ind-F6	F	25	Indonesia	Undergraduate	B1_1	Indonesian	285
ICN-Kor-F1	F	24	Korea	Undergraduate	B2_0	Korean	309
ICN-Pun-F1	F	21	Pakistan	Undergraduate	B1_2	Punjabi	459
ICN-Pun-F2	F	21	Pakistan	Undergraduate	B1_2	Punjabi	516
ICN-Pun-F3	F	20	Pakistan	Undergraduate	B1_2	Punjabi	557
ICN-Pun-M1	M	20	Pakistan	Undergraduate	B1_2	Punjabi	556
ICN-Pun-M2	M	21	Pakistan	Undergraduate	B1_2	Punjabi	703
ICN-Pun-F4	F	24	Pakistan	Undergraduate	B1_2	Punjabi	562
ICN-Tha-F1	F	18	Thailand	Undergraduate	B1_2	Thai	312
ICN-Tha-M1	M	19	Thailand	Undergraduate	B1_2	Thai	265
ICN-Tha-M2	M	18	Thailand	Undergraduate	B1_2	Thai	402
ICN-Tha-M3	M	18	Thailand	Undergraduate	B1_2	Thai	399
ICN-Tha-F2	F	18	Thailand	Undergraduate	B1_2	Thai	391
ICN-Tha-F3	F	20	Thailand	Undergraduate	B1_2	Thai	385
ICN-Tha-F4	F	19	Thailand	Undergraduate	B1_2	Thai	425
ICN-Man-F3	F	18	Taiwan	Undergraduate	A2_0	Mandarin	319
ICN-Man-F4	F	20	Taiwan	Undergraduate	B1_1	Mandarin	310
ICN-Man-F5	F	23	China	Master's	B1_2	Mandarin	2,306
ICN-Man-F6	F	21	China	Undergraduate	B1_1	Mandarin	1,657
ICN-Man-M3	M	22	China	Undergraduate	B2+	Mandarin	2,571
ICN-Man-F7	F	19	China	Undergraduate	B1_1	Mandarin	1,830
ICN-Man-M4	M	19	Hong Kong	Undergraduate	B1_1	Mandarin	2,727
ICN-Can-F1	F	22	Hong Kong	Undergraduate	B1_1	Cantonese	1,950
ICN-Can-M1	M	21	Hong Kong	Undergraduate	B1_1	Cantonese	3,326
ICN-Can-M2	M	20	Hong Kong	Undergraduate	B2+	Cantonese	4,608
ICN-Ind-F7	F	22	Indonesia	Undergraduate	B1_2	Indonesian	1,570
ICN-Ind-F8	F	33	Indonesia	Master's	A2_0	Indonesian	1,851
ICN-Ind-F9	F	24	Indonesia	Master's	B1_1	Indonesian	1,771
ICN-Ind-F10	F	24	Indonesia	Undergraduate	B1_2	Indonesian	2,395
ICN-Ind-F11	F	20	Indonesia	Undergraduate	B1_2	Indonesian	2,453
ICN-Ind-M4	M	25	Indonesia	Master's	B1_2	Indonesian	2,326
ICN-Jap-M1	M	18	Japan	Undergraduate	B1_1	Japanese	1,681
ICN-Jap-F1	F	21	Japan	Undergraduate	B1_2	Japanese	2,431
ICN-Pun-F5	F	21	Pakistan	Master's	B1_1	Punjabi	1,806
ICN-Pun-M3	M	21	Pakistan	Undergraduate	B1_2	Punjabi	1,912
ICN-Pun-M4	M	24	Pakistan	Undergraduate	B1_2	Punjabi	1,694
ICN-Pun-M5	M	22	Pakistan	Master's	B1_1	Punjabi	1,744
ICN-Tha-F5	F	23	Thailand	Undergraduate	B1_2	Thai	2,459
ICN-Tha-F6	F	23	Thailand	Undergraduate	B1_2	Thai	1,628
ICN-Tha-M4	M	22	Thailand	Undergraduate	B1_2	Thai	1,603
ICN-Tha-M5	M	21	Thailand	Undergraduate	B1_1	Thai	5,394
ICN-Tha-M6	M	20	Thailand	Undergraduate	B1_1	Thai	1,998
ICN-Tha-F7	F	21	Thailand	Undergraduate	B1_2	Thai	3,939
ICN-Tha-F8	F	21	Thailand	Undergraduate	B1_2	Thai	2,408

Table 3 (continued)

Speaker Code	Sex	Age	Nationality	Education	Proficiency	L1	No. of words
ICN-Tha-F9	F	20	Thailand	Undergraduate	B1_1	Thai	2,003
ICN-Tha-M7	M	21	Thailand	Undergraduate	B1_2	Thai	2,538
ICN-Tha-F10	F	24	Thailand	Undergraduate	B1_2	Thai	2,367
ICN-Man-M5	M	19	Taiwan	Undergraduate	B1_2	Mandarin	1,286
ICN-Man-M6	M	21	Taiwan	Undergraduate	B2+	Mandarin	2,222

Note. Speaker codes follow the format ICN-[L1]-[sex, sequence number], where L1 codes denote speakers' L1 (e.g., Man = Mandarin), sex is coded as F or M, and sequence numbers are assigned within each L1–sex group. Numerical suffixes in utterance-level identifiers (e.g., ICN-Man-F1-1, ICN-Man-F1-2) indicate multiple corpus files for the same speaker and are used solely for traceability, not to distinguish speakers.

Data Coding

All extracted tokens were coded using a scheme capturing both macro-level (corpus-wide distribution) and micro-level (speaker-specific and contextual) features. Macro-level coding enabled quantitative analysis of overall frequency patterns across demographic and corpus variables; micro-level coding captured variation in speakers' contextualised deployment and the discourse functions of each construction. Coding was carried out systematically in Excel to ensure consistency and reproducibility across ICE, ACE, and ICNALE.

Quantitative Variables

For each speaker, the following variables were recorded:

- Demographics: sex, age, L1, education, and (for learner data) proficiency level.
- Total spoken word count
- Frequencies:
 - *it's* + *V*Ø and *that's* + *V*Ø
 - occurrences with intervening elements
 - canonical *it's*, and *that's*
- Pronoun–verb patterns:
 - singular pronouns (*he, she, it, this, that*) + bare/uninflected verbs (referred to as *pronoun–bare verb*)
 - singular pronouns (*he, she, it, this, that*) + –s inflected verbs (referred to as –s *inflected pronoun–verb*)
- Register or interactional setting for each token

Ambiguous cases (e.g., unclear boundaries, transcription uncertainty) were checked against source files and verified for accuracy.

Qualitative Variables

Each token was additionally coded for functional and discourse-level features, including:

- pragmatic or functional role of intervening elements
- indicators of cognitive entrenchment (e.g., repetition, priming)
- intra-speaker alternation between standard and nonstandard forms

These variables allowed analysis of how speakers use the constructions in context, complementing quantitative distributional patterns with functional and cognitive insights.

Coding Reliability

The coding scheme was piloted on sample data from each corpus, refined iteratively, and applied consistently across datasets. A subset of tokens was double-coded at different times to check internal stability. Although independent second-coder validation was not feasible, the procedures ensured reliable and transparent coding.

Quantitative Analysis

This study employs an integrated quantitative–qualitative approach to examine the distribution, associations, and discourse functions of *it's + Vø* and *that's + Vø* across ICE, ACE, and ICNALE. Four complementary analyses were conducted: (i) frequency and distribution analysis (to examine variation across corpora and speaker variables), (ii) collocational analysis (to identify verb co-occurrence patterns), (iii) collostructional analysis (to assess verb–construction associations), and (iv) correlation analysis (to test relationships with related pronoun–bare verb patterns).

Frequency and Distribution

Raw and normalised frequencies were calculated for demographic and corpus variables (register, sex, age, L1, education, and learner proficiency). Corpus-level frequencies were normalised per million words (pmw) to enable comparison across datasets. Variation was tested using two-tailed independent *t*-tests and one-way ANOVA.

Collocational Analysis

A five-word rightward window identified verb lemmas co-occurring with *it's* and *that's*. Measures included:

- verb distribution and distance from the node;
- intervening element types and functions; and
- association strength using Pointwise Mutual Information (PMI; Church & Hanks, 1990).

PMI was selected for its sensitivity to emergent patterning. Values above 2 were interpreted as indicating meaningful associations, with positive PMI values signalling a stronger attraction to *it's + Vø* and negative values indicating a stronger attraction to *that's + Vø*.

Collostruction Analysis

To compare verb preferences between the two constructions, 2×2 contingency tables were generated for each verb. Odds Ratio (OR) measured direction of preference (OR > 1 = favours *it's + Vø*; OR < 1 = favours *that's + Vø*), and Fisher's Exact Test provided exact significance values. Collostructional strength was computed as $-\log_{10}(p)$. Values above 1 (approximately corresponding to $p < .05$) indicate substantive attraction or repulsion.

Correlation Analysis

Spearman's rho tested relationships between constructional usage and related *pronoun–bare verb* patterns, including:

- emergent vs. canonical forms;
- cross-constructional relations with *pronoun-bare verb* patterns;
- sequential proximity to canonical *it's*, and *that's*; and
- parallel *pronoun-bare verb* patterns before and after each construction.

Qualitative Analysis

This analysis illuminates how speakers deploy the constructions interactionally, focusing on their discourse functions, pragmatic roles, and contribution to discourse organisation and communicative alignment.

RESULTS AND DISCUSSION

Overview

This section presents the findings in relation to the four research objectives, integrating quantitative and qualitative evidence to account for the emergent *it's + V $\emptyset$* and *that's + V $\emptyset$* constructions. It first establishes cross-corpus differences in frequency and distribution across institutionalised varieties, ELF interactions, and learner Englishes, and then examines the structural, functional, and cognitive motivations underlying these patterns through collocational, collostructional, and discourse-based evidence. It then explores how usage contexts shape constructional behaviour, drawing on correlation patterns and intra-speaker variation. Finally, the findings are synthesised to demonstrate how high-frequency copular templates function as productive resources for grammatical innovation within usage-based and constructional frameworks.

Objective 1: Comparative distribution across institutionalised, ELF, and learner varieties

Frequency patterns reveal clear cross-corpus contrasts in the distribution of *it's + V $\emptyset$* and *that's + V $\emptyset$* . *It's + V $\emptyset$* is substantially more frequent in ICNALE (86 tokens; 40.95 pmw) and ACE (29 tokens; 29.00 pmw) than in ICE (7 tokens; 3.89 pmw), indicating greater productivity in learner and ELF contexts. In contrast, *that's + V $\emptyset$* remains consistently rare across all corpora (7–12 tokens; 3.89–12.00 pmw), suggesting limited diffusion.

This asymmetry indicates that multilingual and acquisitional contexts facilitate the expansion of *it's + V $\emptyset$* , while *that's + V $\emptyset$* remains comparatively restricted. The pattern aligns with usage-based expectations that high-frequency chunks such as *it's* provide a stronger analogical base for extension than less frequent alternatives.

Across registers, *it's + V $\emptyset$* clusters overwhelmingly in informal contexts (2.78–40.95 pmw), whereas *that's + V $\emptyset$* remains low and sporadic in both formal and informal settings. However, no significant register effects emerge ($p > .16$), indicating that situational formality does not strongly constrain distribution. Instead, the spread of *it's + V $\emptyset$* appears to operate below the level of conscious stylistic control.

Sociodemographic patterns further support its emergent status. In ACE and ICNALE, female speakers use *it's + V $\emptyset$* more frequently, with a significant effect in ICNALE ($t(54) = 2.04$, $p = .046$, $d = 0.56$). Younger speakers also favour the construction, with a significant difference between <20 and 20–30 groups in ICNALE ($t(53) = 2.06$, $p = .045$, $d \approx 0.74$). These patterns suggest that *it's + V $\emptyset$* is advancing through socially less constrained groups typically associated with linguistic innovation.

L1 background plays a role in ELF contexts (ACE: $F(4,13) = 5.42$, $p = .009$), with Burmese, Thai, and Malaysian Malay speakers showing higher usage, whereas no significant L1 effects appear in ICNALE ($p > .20$), likely reflecting task-controlled production. Education shows no significant effects ($p > .05$), and proficiency differences in ICNALE are non-significant ($F(3,51) = 0.68$, $p = .568$), although mid-level learners (B1_2) show a descriptive peak, suggesting a possible zone of heightened analogical experimentation.

Collectively, these results demonstrate that *it's + Vø* is most productive in ELF and learner varieties, where interactional flexibility and reduced normative pressure enable the expansion of emergent constructions.

While these distributional patterns establish where *it's + Vø* and *that's + Vø* occur and how widely they diffuse across varieties, they do not yet explain how these constructions are internally organised or why one expands more readily than the other. To account for this asymmetry, the analysis now turns to the structural, functional, and cognitive mechanisms underlying the constructions.

Objective 2: Structural, functional, and cognitive motivations underlying the constructions

Structural evidence from collocational and collostructional analyses shows that *it's + Vø* is undergoing schematic expansion, while *that's + Vø* remains lexically constrained.

Lexical diversity increases markedly from ICE (8 verbs) to ACE (17 verbs) and ICNALE (27 verbs), indicating progressive expansion of the verb slot (for a complete list of verbs attested in *it's + Vø* and *that's + Vø* across each corpus, see Appendix D). Recurrent cross-varietal pivots such as *mean*, *depend*, and *make* illustrate how specific verbs anchor early stages of constructional growth. In contrast, *that's + Vø* exhibits narrower lexical distribution, suggesting more limited schematic productivity.

Collocational strength further supports this asymmetry (Table 4). PMI results show that *it's + Vø* maintains moderate associations across multiple verbs, consistent with an emerging schematic construction. By contrast, *that's + Vø* displays stronger but highly selective associations (e.g., *mean* in ICE; *go* in ACE and ICNALE). Collostructional analysis reinforces this pattern: in ICE, *mean* significantly favours *that's + Vø* (strength = 1.39, $p = .041$), while in ICNALE, *go* shows a strong association (strength = 2.09, $p = .008$). Other verbs, such as *have*, *help*, *look*, and *make*, exhibit weak, non-significant tendencies (PMI values -1 to $+1.9$), indicating that *it's + Vø* expands through diffuse, low-strength associations rather than strong lexical fixation (for a complete list of verb-by-verb PMI scores, see Tables E1–E3 in Appendix E).

From a cognitive perspective, these patterns reflect schema formation through analogical extension. The construction is not generated through rule-based processes but emerges from repeated usage events, where low-level associations gradually generalise into a productive template. The asymmetry between *it's + Vø* and *that's + Vø* therefore reflects differences in entrenchment and accessibility of their source constructions.

However, these structural tendencies alone do not determine how the constructions are realised in actual discourse. Their development and distribution are also shaped by the contexts in which they are used.

Table 4. Cross-Corpus PMI Score Comparison for Recurring Verbs

<i>V</i> ∅	ICE	ACE	ICNALE	Cross-Corpus Pattern
<i>come</i>	—	+0.38	-0.80	Weak association with <i>it's</i> + <i>V</i> ∅ in ACE and <i>that's</i> + <i>V</i> ∅ in ICNALE
<i>depend</i>	—	+2.70	+1.40	Consistently favours <i>it's</i> + <i>V</i> ∅ in both ACE and ICNALE
<i>give</i>	+0.36	-7.65	+0.34	Weak association in ICE and ICNALE; Strong preference for <i>that's</i> + <i>V</i> ∅ in ACE
<i>go</i>	—	-8.64	-11.72	Exclusively favours <i>that's</i> + <i>V</i> ∅ in ACE and ICNALE
<i>have</i>	—	+0.38	-0.93	Weak preference for <i>it's</i> + <i>V</i> ∅ in ACE and <i>that's</i> + <i>V</i> ∅ in ICNALE
<i>help</i>	—	+0.38	-1.26	Weak association with <i>it's</i> + <i>V</i> ∅ in ACE; Mild preference for <i>that's</i> + <i>V</i> ∅ in ICNALE
<i>look</i>	+1.91	+0.38	—	Mildly favours <i>it's</i> + <i>V</i> ∅ in ICE and ACE
<i>make</i>	—	-1.27	-1.26	Mildly favours <i>that's</i> + <i>V</i> ∅ in ACE and ICNALE
<i>mean</i>	-2.32	+0.46	-0.80	Weak association in ACE and ICNALE; Mild preference for <i>that's</i> + <i>V</i> ∅ in ICE.
<i>take</i>	+1.91	-1.27	-0.80	Weak association in ICNALE; Mild preference for <i>it's</i> + <i>V</i> ∅ in ICE and <i>that's</i> + <i>V</i> ∅ in ACE.

Note. A dash (—) indicates that no tokens were attested.

Objective 3: The role of usage contexts in shaping constructional behaviour

Correlation analyses reveal distinct contextual dynamics across corpora. In ICE, canonical *it's* clusters locally ($p = .007$), but neither *it's* + *V*∅ nor *that's* + *V*∅ correlates with canonical variables, suggesting early-stage emergence with limited discourse integration. In ACE, stronger clustering patterns emerge: canonical *it's* and pronoun-bare verb sequences co-occur significantly with *it's* + *V*∅ ($p < .001$), while *that's* + *V*∅ also clusters with canonical and non-canonical sequences ($p = .020$; $p = .034$). ICNALE shows similarly robust clustering ($p < .001$), indicating reliance on formulaic sequencing in learner production. Collectively, these patterns suggest that ELF and learner contexts promote discourse chaining and local repetition, facilitating the spread of emergent constructions through sequential co-occurrence rather than global grammatical integration.

Qualitative evidence further clarifies how these mechanisms interact. Node-verb patterns (Table 5) show that while both constructions favour adjacency, *it's* + *V*∅ exhibits greater flexibility in ICNALE, where interveners such as *really* and *not* frequently occur. This suggests that speakers are actively testing constructional slots during early stages of entrenchment.

Table 5. Intervening Elements across ICE, ACE, and ICNALE

Corpus	Node	Intervening Elements	N = 36		<i>V</i> ∅
			<i>n</i>	%	
ICE	<i>it's</i>	<i>almost</i>	1	2.78	<i>remain</i>
		<i>fully</i>	1	2.78	<i>depend</i>
		<i>only</i>	1	2.78	<i>have</i>
ACE	<i>it's</i>	<i>more</i>	1	2.78	<i>guarantee</i>
		<i>really</i>	1	2.78	<i>depend</i>
		<i>not</i>	1	2.78	<i>depend</i>

Table 5 (continued)

Corpus	Node	Intervening Elements	N = 36		Vø
			n	%	
ICNALE	it's	<i>totally</i>	1	2.78	<i>affect</i>
		<i>very</i>	2	5.56	<i>bother, destroy</i>
		<i>more</i>	1	2.78	<i>need</i>
		<i>just</i>	3	8.33	<i>cause, give, harm</i>
		<i>really^a</i>	5	13.89	<i>create, disturb, give, have, help</i>
		<i>also</i>	5	13.89	<i>damage, destroy, have, help</i>
		<i>not</i>	6	16.67	<i>bring, give, influence, take</i>
		<i>not only</i>	6	16.67	<i>affect, damage, do, make</i>
		<i>that's mentally</i>	1	2.78	<i>affect</i>

Note. Percentages are based on N = 36 and are provided for descriptive purposes only.

^a Includes one use of *real* (adjective) for *really* (adverb) by speaker ICN-Ind-M2 (*it's real give them skill*).

Table 6. Categorisation of Intervening Elements by Functional Group

Category	Intervening Elements	N = 36	
		n	%
Degree/Intensity	<i>almost, fully, just, more, really, totally, very</i>	16	44.44
Negator/Contrastive	<i>not, not only</i>	13	36.11
Additive/Conjunctive	<i>also</i>	5	13.89
Restrictive	<i>only</i>	1	2.78
Manner/Evaluative	<i>mentally</i>	1	2.78

Note. Percentages are based on N = 36 and are provided for descriptive purposes only.

Table 6 shows that intervening elements are concentrated in a small set of functional categories, dominated by degree/intensity (44.44%) and negation/contrast (36.11%), with additive, restrictive, and evaluative functions occurring marginally. These elements function primarily **as stance markers**, as illustrated below:

1. *It is – **it's just give** harms – smoking just gives harm* (ICN-Pun-F1)
2. *You can take it to anywhere and - and **it's really help** you to study and to make a lot of things.* (ICN-Man-M3-2)
3. *that's your decision **it's not depend** on the contract at all* (ACE-MY-VieF1)
4. *it's a very bad manner in the restaurant, yeah, and, um – uh – it's – **it's not only damage** your health* (ICN-Tha-F4)
5. *the government take action on it because smoking is a bad and it's – **it's totally affect** the environment* (ICN-Pun-F3)
6. *when my friend and I were having a birthday celebration in a restaurant, uh, we noticed that some - some - there is a guy - there was a guy smoking near - near us and that has, uh – **that's mentally effect** our appetite.* (ICN-Man-M6)

These examples show that intensifiers and negators enable speakers to express evaluation, emphasis, and contrast within a flexible constructional frame. Crucially, they extend from canonical templates (e.g., *it's really + ADJ*, *it's not + X*) into *it's + Vø* environments, indicating analogical projection from entrenched copular constructions rather than rule-based generation.

Intra-speaker variation provides further evidence of contextual adaptation. In ICE, alternation between *it's* + *V*Ø, *that's* + *V*Ø, and canonical forms is infrequent and unsystematic, suggesting limited productive flexibility:

ICE-HK-TagF1 (*mean*):

1. So **that's mean** you just uh <,> have uh
2. So <,> in **that means** <,> you mean that you got uh some ideas from her
3. You know what it is, kin **that means** that you have close

ICE-HK-CanM1 (*mean*):

1. So **that's mean** we do not have enough <?> spared </?> classroom to support all the classes
2. **That means** <,> some of the classes will not will not <,> uh have any home room
3. When I when I say a <?> free </?> lesson **that means** a lesson will
4. So if we have more this sort of multi period lessons **that means** uh we we do not have <,> uh
5. Sorry five uh twenty five percent of uh schools use computers only for constraint checking <,> **that means**, uh, they have to make a decision
6. **That means**, uh if you if a school has a one hundred
7. So that **means** there are a lot of periods to be filled up, okay, but manually

In ACE, speakers (e.g., ACE-MY-MalM1) show more dynamic alternation across constructions within discourse:

1. S1: i didn't feel jealous i mean <pvc>lah</pvc>i didn't (.) **it's mean** i didn't develop
2. S1: yeah (.) **it's mean** the lecturer didn't notice me, **he noticed** someone else
3. S1: it's not it's not develop **[it's mean]** that you:: do your responsible <coughs> -lity
4. S1: i already do the er the photocopy and then **[she ask]** me er she said she told me like that
5. S1: Can you call [first name8] call [first name8] i don't think so **[he already picked]**
6. S1: and then even with [first name3] also i didn't speak at all **[he always looks]** down on me

This reflects a more strategic deployment of available forms for pragmatic alignment and discourse management.

ICNALE shows the highest degree of variability, with extensive alternation across constructions and verbs.

ICN-Man-F5:

1. The school's systems also use, uh, Windows, so we have no choice but to, uh, touch the MacBook or like that. So, **[it's need]** to take times about 2 or 3 days that - uh, to use - to use about, uh, the Mac - the Mac. (ICN-Man-F5-2)
2. Like I am a foreigner, uh, student, so **[it takes]** a lot of money, uh, for me if I want to continue my study in Japan, and as you know, the salary in Japan is totally higher than China, so **[it's make]** me feel a little successful. (ICN-Man-F5-3)
3. uh, I feel that I am a new people. Yes, **[that's make]** me feel exciting. (ICN-Man-F5-4)
4. this restaurant near Osaka station that's called Dancing Club. **That's have** seafood and my friend is just that day have a, uh, birthday... We can smell that - uh, everyone can smell that, so, um, **it's influence** us, so I think maybe it should be, uh, abandoned. (ICN-Man-F5-6)

5. *My friend, he – **she has** a sick, so she can't - uh, she can't stay there anymore* (ICN-Man-F5-5)
6. *Eddie was, uh, thinking about that maybe he can going to surfer, uh, when the day is sunny, but when **he consider** about his pocket, he - he thought that **he have** no money to go out - go to travel.* (ICN-Man-F5-1)

ICN-Tha-M5:

1. *Actually, I have a quote, that I - that's the reason why I love languages, which is like from Nelson Mandela, which is like what - what does it say, **it say** like if you talk to a man in the language he understands, **that's go** to his head, but if you talk to the man in his language, **that's go** to his heart. So yeah, so that exactly---* (ICN-Tha-M5-1)
2. *So, **he** basically **goes to that shop and tries to make- tries to- like, tries** to interview in order to get a job. Now, he got a work, and also ask for mom, like, how to say, **he get** salary, and **he** just **try** to keep some. And then so, finally after 4 months passed, **he collect** the money enough in order to buy a notebook, what he wants, wanted earlier. **He go** to work regularly - regularly, and tried to ask his - his mother for the money, yes.* (ICN-Tha-M5-2)

These patterns indicate ongoing analogical experimentation under reduced grammatical constraint, where speakers flexibly recruit available templates to meet communicative demands.

Overall, usage context—particularly ELF interaction and learner production—facilitates flexibility, repetition, and analogical extension, shaping constructional behaviour in ways that differ from more stabilised varieties. These context-sensitive patterns point to a dynamic interaction between structure, usage, and discourse, motivating a broader account of how such constructions emerge and stabilise within the grammar.

Objective 4: Implications for usage-based and constructional accounts of grammatical innovation

This study makes three contributions to usage-based and constructional accounts of grammatical innovation. First, it demonstrates how high-frequency subject-copula units (*it's*, *that's*) function as productive templates for emergent syntactic patterning in multilingual and acquisitional Englishes. Second, it shows that *it's* + *V*Ø and *that's* + *V*Ø follow distinct trajectories of entrenchment, shaped by differences in frequency and distributional range. Third, it provides empirical evidence that grammatical innovation is jointly driven by frequency, analogy, and interactional pressures.

Taken together, the study provides empirical evidence that grammatical structure emerges from usage, showing how conventionalised forms can be repurposed as flexible templates for new constructions. In doing so, it extends usage-based and constructional models by showing how the interaction of frequency, analogy, and discourse across diverse Englishes shapes innovation.

CONCLUSION

This study investigated the emergent constructions *it's* + *V*Ø and *that's* + *V*Ø across institutionalised Asian Englishes, ELF interaction, and learner Englishes, showing that their development follows systematic, usage-driven pathways rather than representing random deviation from canonical copular structures. Across the four objectives, the findings

converge on a single pattern of asymmetrical constructional emergence shaped by frequency, analogy, and discourse use in multilingual ecologies.

First, the comparative analysis shows that *it's* + *V*ø is widely diffused in ELF and learner contexts, whereas *that's* + *V*ø remains comparatively restricted across varieties. This asymmetry suggests that high-frequency forms such as *it's* provide a stronger base for analogical extension and constructional expansion than less frequent counterparts.

Second, the structural and cognitive evidence indicates that *it's* + *V*ø is undergoing schematic broadening through diffuse lexical associations and analogical projection from entrenched copular patterns, while *that's* + *V*ø exhibits more selective verb alignment and constrained productivity. These differences reflect uneven entrenchment and differential access to productive templates in usage.

Third, usage context plays a central role in shaping constructional behaviour. ELF and learner settings promote repetition, discourse chaining, and flexible alternation between canonical and non-canonical patterns. These environments facilitate the real-time adaptation of constructions for interactional purposes such as evaluation, emphasis, and alignment, highlighting the role of discourse conditions in driving grammatical innovation.

Taken together, the findings demonstrate that high-frequency subject-copula units function as productive scaffolds for emergent syntactic patterns. This study therefore contributes to usage-based and constructional accounts of grammatical change by showing that grammatical innovation arises from the interaction of frequency, analogy, and discourse-pragmatic function, rather than rule-based extension alone. Crucially, the results challenge deficit-oriented interpretations of non-canonical patterns in Asian Englishes. Instead, they support the view that multilingual Englishes are sites of systematic grammatical innovation shaped by dense language contact, variable normative pressures, and evolving communicative needs.

The comparative design of this study necessarily entails limitations arising from the corpora used. Differences in compilation principles, demographic distributions, register coverage, and data collection contexts across ICE, ACE, and ICNALE constrain direct comparability and may influence frequency and functional patterns. Each corpus also presents specific limitations, including variation in collection periods (ICE), uneven site representation and missing file dates (ACE), and the task-based, topic-restricted nature of ICNALE, which reduces interactional spontaneity. Occasional absences of canonical *it's*, or *that's* near emergent tokens, likely reflect these design features rather than speaker behaviour. In addition, manual coding introduces an element of interpretive judgement. Although reliability procedures were applied, the absence of a second independent coder and the context-sensitive nature of the constructions mean that some classifications may vary slightly across analysts. These limitations do not undermine the overall findings but should be considered when interpreting fine-grained patterns.

Future research should extend this line of inquiry through larger and more regionally diverse corpora to better capture the ecological conditions that promote or inhibit constructional diffusion. Longitudinal ELF and learner datasets would enable tracking of developmental trajectories over time, while experimental and processing-based studies could further clarify the cognitive mechanisms underlying analogical extension and entrenchment. Greater attention to prosody, discourse organisation, and sequential positioning may also reveal how these constructions function as interactional resources

beyond their structural properties. Finally, future work should investigate whether similar template-based innovations emerge around other high-frequency copular units, contributing to a broader understanding of constructional change across Asian Englishes.

REFERENCES

- Asian Corpus of English (ACE). (2024). *The Asian Corpus of English*. <https://corpus.eduhk.hk/ace/index.html>
- Bernaish, T., Mendis, D. & Mukherjee, J. (2019). *Manual to the International Corpus of English – Sri Lanka*. Giessen: Justus Liebig University, Department of English.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Brezina, V., Weill-Tessier, P., & McEnery, A. (2021). LancsBox 6.0. Lancaster University.
- Bybee, J. & Scheibman, J. (1999). The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics*, 37(4), 575–596. <https://doi.org/10.1515/ling.37.4.575>
- Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22–29.
- Díaz-Campos, M. & Pollock, M. (2023). The Future of Usage-Based Sociolinguistics. In M. Díaz-Campos & S. Balasch (Eds.), *The Handbook of Usage-Based Linguistics* (pp. 509–526). Hoboken, NJ: Wiley-Blackwell.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. <https://doi.org/10.1017/S0272263102002024>
- Ellis, N. C. (2017). Chunking in Language Usage, Learning and Change: I Don't Know. In M. Hundt, S. Mollin, & S. E. Pfenninger (Eds.), *The Changing English Language: Psycholinguistic Perspectives* (pp. 113–147). Cambridge: Cambridge University Press.
- Ellis, R. (1997). *Second Language Acquisition*. Oxford: Oxford University Press.
- Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Greenbaum, S. (1996). Introducing ICE. In S. Greenbaum (Ed.), *Comparing English worldwide: The International Corpus of English* (pp. 3–12). Oxford: Oxford University Press.
- Hamid, M. O., & Hasan, M. M. (2020). Bangladeshi English. In B. Kingsley, W. Botha, & A. Kirkpatrick (Eds.), *The Handbook of Asian Englishes* (pp. 297–315). Hoboken, NJ: Wiley-Blackwell.
- International Corpus of English (ICE). (n.d.). *International Corpus of English*. University of Zurich. <https://www.ice-corpora.uzh.ch/en.html>
- Ishikawa, S. (2023). *The ICNALE Guide: An Introduction to a Learner Corpus Study on Asian Learners' L2 English*. London: Routledge.
- Levshina, N., & Lorenz, D. (2022). Communicative Efficiency and the Principle of No Synonymy: Predictability effects and the variation of *want to* and *wanna*. *Language and Cognition*, 14(2), 249–274. <https://doi.org/10.1017/langcog.2022.7>
- Mauranen, A. (2012). *Exploring ELF*. Cambridge University Press.

- Mauranen, A. (2018). Conceptualising ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge Handbook of English as a Lingua Franca* (pp. 7–24). London: Routledge.
- Mohd Yusoff, Y., Salehuddin, K., Abdullah, I. H., & Toran, H. (2019). English Morphosyntactic Performance of a High-Functioning ASD Child: Implications for ELT. *Malaysian Journal of Learning and Instruction*, 16(1), 155–179. <https://doi.org/10.32890/mjli2019.16.1.7>
- Mufwene, S. S. (2001). *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.
- Nelson, G. (2002). *The ICE India Corpus* [User Manual]. International Corpus of English. <https://www.ice-corpora.uzh.ch/en.html>
- Nelson, G. (2006). *The ICE Hong Kong Corpus* [User Manual]. International Corpus of English. <https://www.ice-corpora.uzh.ch/en.html>
- Poulisse, N. (1999). *Slips of the Tongue: Speech Errors in First and Second Language Production*. Amsterdam: John Benjamins.
- Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.
- Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics*, 24, 209–239. <https://doi.org/10.1017/S0267190504000145>
- Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55(2), 157–166. <https://doi.org/10.1016/j.jml.2006.03.007>
- Vetchinnikova, S. & Hiltunen, T. (2021). ELF and Language Change at the Individual Level. In A. Mauranen & S. Vetchinnikova (Eds.), *Language Change: The Impact of English as a Lingua Franca* (pp. 205–233). Cambridge: Cambridge University Press.
- Wang, L., & Kirkpatrick, A. (2024). The Asian Corpus of English (ACE). In A. J. Moody (Ed.), *The Oxford Handbook of Southeast Asian Englishes* (pp. 753–767). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780192855282.013.42>
- Zayani, Z.A. (2024). 'It's Depend', 'Its Depend', or 'It Depends'? A Portrait of ELF Speaking Variations. *J-Lalite: Journal of English Studies*, 5(1), 1–19. <https://doi.org/10.20884/1.jes.2024.5.1.11188>

Appendix A

Table A. Standard Design of Spoken Component of ICE

Spoken (300)	Domains	Text categories	Filenames
Dialogues (180)	Private (100)	Face-to-face conversations (90)	S1A-001 to S1A-090
		Phone Calls (10)	S1A-091 to S1A-100
	Public (80)	Classroom Lessons (20)	S1B-001 to S1B-020
		Broadcast Discussions (20)	S1B-021 to S1B-040
		Broadcast Interviews (10)	S1B-041 to S1B-050
		Parliamentary Debates (10)	S1B-051 to S1B-060
		Legal cross-examinations (10)	S1B-061 to S1B-070
		Business Transactions (10)	S1B-071 to S1B-080
		Spontaneous commentaries (20)	S2A-001 to S2A-020
		Unscripted Speeches (30)	S2A-021 to S2A-050
Monologues (120)	Unscripted (70)	Demonstrations (10)	S2A-051 to S2A-060
		Legal Presentations (10)	S2A-061 to S2A-070
		Broadcast News (20)	S2B-001 to S2B-020
	Scripted (50)	Broadcast Talks (20)	S2B-021 to S2B-040
		Non-broadcast Talks (10)	S2B-041 to S2B-050

Note. Numbers in brackets indicate the number of files.

Appendix B

Table B. Selected ACE Dataset

Filename	Speaker Code
BN_ED_wsd_preparing caseline_facebook versus face-to-face interactions	ACE-BN-MalM1
CN_PO_int_china-saudi arabia bilateral relationship	ACE-CN-ChiF1
HK_PRS_qas_surgical technique	ACE-HK-CanM1
MS_ED_con_1	ACE-MY-ThaF1
MS_ED_con_2	ACE-MY-ThaF2
MS_ED_con_3	ACE-MY-MalM1
MS_ED_con_4	ACE-MY-MalM2
MS_ED_con_6	ACE-MY-ManF1
MS_ED_con_12	ACE-MY-BurF1
MS_LE_con_5	ACE-MY-BurF2
MS_LE_con_6	ACE-MY-ThaF3
MS_LE_con_11	ACE-MY-VieF1
MS_LE_con_11	ACE-MY-KhmM1
MS_PB_con_1	ACE-MY-VieF2
MS_PB_con_6	ACE-MY-ThaM1
PH_ED_sed_recounting lesson	ACE-PH-VieF1
PH_ED_sed_recounting lesson	ACE-PH-BurF1
SG_ED_con_4	ACE-SG-MalF1
SG_ED_con_5	ACE-SG-ManF1
VN_ED_qas_tesol_localised teaching strategy	ACE-VN-VieF1
VN_LE_con_culture	ACE-VN-VieF2
VN_LE_con_pho restaurant	ACE-VN-VieF3
VN_LE_con_pho restaurant	ACE-VN-VieF4
VN_LE_con_jobs and professions 2	ACE-VN-VieF5

Note. Country codes: BN = Brunei; CN = Mainland China; HK = Hong Kong; JP = Japan; MS = Malaysia; PH = Philippines; SG = Singapore; TW = Taiwan; VN = Vietnam. Setting types: ED = education; LE = leisure; PB = professional business; PO = professional organisation; PRS = professional research/science. Speech event types: int = interview; prc = press conference; sed = seminar discussion; wgd = working group discussion; wsd = workshop discussion; mtg = meeting; qas = question-answer session; con = conversation.

Appendix C

Table C. Selected ICNALE Dataset

Filename	Speaker Code	Filename	Speaker Code
SM_CHN_SMK1_022_B1_2	ICN-Man-M1	SD_HKG_001_12_SMK_RP_Main	ICN-Man-M4
SM_CHN_SMK1_045_B1_2	ICN-Man-M2	SD_HKG_009_04_PTJ_PD_QA	ICN-Can-F1-1
SM_CHN_SMK1_102_B1_2	ICN-Man-F1-1	SD_HKG_009_14_Reflection	ICN-Can-F1-2
SM_CHN_SMK2_102_B1_2	ICN-Man-F1-2	SD_HKG_009_12_SMK_RP_Main	ICN-Can-F1-3
SM_CHN_SMK1_113_A2_0	ICN-Man-F2	SD_HKG_017_01_Intro	ICN-Can-M1
SM_IDN_PTJ2_006_B1_2	ICN-Ind-M1	SD_HKG_025_06_PTJ_RP_Main	ICN-Can-M2
SM_IDN_PTJ1_050_B1_1	ICN-Ind-M2	SD_IDN_016_07_PTJ_RP_QA	ICN-Ind-F7-1
SM_IDN_PTJ2_064_A2_0	ICN-Ind-F1	SD_IDN_016_09_SMK_PD_Main	ICN-Ind-F7-2
SM_IDN_SMK1_025_B1_1	ICN-Ind-F2-1	SD_IDN_017_14_Reflection	ICN-Ind-F8
SM_IDN_SMK2_025_B1_1	ICN-Ind-F2-2	SD_IDN_019_07_PTJ_RP_QA	ICN-Ind-F9
SM_IDN_SMK2_019_B1_1	ICN-Ind-M3	SD_IDN_023_01_Intro	ICN-Ind-F10
SM_IDN_SMK2_056_A2_0	ICN-Ind-F3	SD_IDN_025_13_SMK_RP_QA	ICN-Ind-F11-1
SM_IDN_SMK2_070_A2_0	ICN-Ind-F4	SD_IDN_025_01_Intro	ICN-Ind-F11-2
SM_IDN_SMK2_072_B1_2	ICN-Ind-F5	SD_IDN_027_01_Intro	ICN-Ind-M4-1
SM_IDN_SMK2_076_B1_1	ICN-Ind-F6	SD_IDN_027_10_SMK_PD_QA	ICN-Ind-M4-2
SM_KOR_SMK1_067_B2_0	ICN-Kor-F1	SD_JPN_039_07_PTJ_RP_QA	ICN-Jap-M1
SM_PAK_SMK2_026_B1_2	ICN-Pun-F1	SD_JPN_073_06_PTJ_RP_Main	ICN-Jap-F1
SM_PAK_PTJ1_030_B1_2	ICN-Pun-F2	SD_PAK_009_04_PTJ_PD_QA	ICN-Pun-F5
SM_PAK_SMK2_044_B1_2	ICN-Pun-F3	SD_PAK_011_10_SMK_PD_QA	ICN-Pun-M3
SM_PAK_SMK2_054_B1_2	ICN-Pun-M1	SD_PAK_016_14_Reflection	ICN-Pun-M4
SM_PAK_PTJ1_071_B1_2	ICN-Pun-M2	SD_PAK_023_12_SMK_RP_Main	ICN-Pun-M5
SM_PAK_SMK1_086_B1_2	ICN-Pun-F4	SD_THA_003_13_SMK_RP_QA	ICN-Tha-F5-1
SM_THA_PTJ2_021_B1_2	ICN-Tha-F1	SD_THA_003_14_Reflection	ICN-Tha-F5-2
SM_THA_PTJ1_028_B1_2	ICN-Tha-M1	SD_THA_004_01_Intro	ICN-Tha-F6-1
SM_THA_SMK1_003_B1_2	ICN-Tha-M2	SD_THA_004_04_PTJ_PD_QA	ICN-Tha-F6-2
SM_THA_SMK1_004_B1_2	ICN-Tha-M3	SD_THA_004_07_PTJ_RP_QA	ICN-Tha-F6-3
SM_THA_SMK1_010_B1_2	ICN-Tha-F2-1	SD_THA_004_14_Reflection	ICN-Tha-F6-4
SM_THA_SMK2_010_B1_2	ICN-Tha-F2-2	SD_THA_019_13_SMK_RP_QA	ICN-Tha-M4
SM_THA_SMK2_019_B1_2	ICN-Tha-F3	SD_THA_021_01_Intro	ICN-Tha-M5-1
SM_THA_SMK2_043_B1_2	ICN-Tha-F4	SD_THA_021_03_PTJ_PD_Main	ICN-Tha-M5-2
SM_TWN_SMK1_013_A2_0	ICN-Man-F3	SD_THA_022_04_PTJ_PD_QA	ICN-Tha-M6
SM_TWN_SMK1_094_B1_1	ICN-Man-F4	SD_THA_024_07_PTJ_RP_QA	ICN-Tha-F7-1
SD_CHN_010_03_PTJ_PD_Main	ICN-Man-F5-1	SD_THA_024_09_SMK_PD_Main	ICN-Tha-F7-2
SD_CHN_010_04_PTJ_PD_QA	ICN-Man-F5-2	SD_THA_027_07_PTJ_RP_QA	ICN-Tha-F8
SD_CHN_010_06_PTJ_RP_Main	ICN-Man-F5-3	SD_THA_033_01_Intro	ICN-Tha-F9-1
SD_CHN_010_10_SMK_PD_QA	ICN-Man-F5-4	SD_THA_033_13_SMK_RP_QA	ICN-Tha-F9-2
SD_CHN_010_12_SMK_RP_Main	ICN-Man-F5-5	SD_THA_033_14_Reflection	ICN-Tha-F9-3
SD_CHN_010_13_SMK_RP_QA	ICN-Man-F5-6	SD_THA_034_10_SMK_PD_QA	ICN-Tha-M7
SD_CHN_039_14_Reflection	ICN-Man-F6	SD_THA_038_01_Intro	ICN-Tha-F10
SD_CHN_045_13_SMK_RP_QA	ICN-Man-M3-1	SD_TWN_029_04_PTJ_PD_QA	ICN-Man-M5-1
SD_CHN_045_04_PTJ_PD_QA	ICN-Man-M3-2	SD_TWN_029_14_Reflection	ICN-Man-M5-2
SD_CHN_047_06_PTJ_RP_Main	ICN-Man-F7	SD_TWN_040_12_SMK_RP_Main	ICN-Man-M6

Note. Numerical suffixes in utterance-level identifiers (e.g., ICN-Man-F1-1, ICN-Man-F1-2) indicate multiple corpus files for the same speaker and are used solely for traceability, not to distinguish speakers.

Appendix D

Table D1a. Verbs Attested in It's + Vø in ICE

<i>Vø</i>	<i>it's + Vø</i>	
	<i>n</i>	Norm.
<i>coincide</i>	1	0.56
<i>continue</i>	1	0.56
<i>give</i>	1	0.56
<i>look</i>	1	0.56
<i>mean</i>	1	0.56
<i>remain</i>	1	0.56
<i>take</i>	1	0.56
Total (<i>N</i>)	7	3.89

Table D1b. Verbs Attested in That's + Vø in ICE

<i>Vø</i>	<i>that's + Vø</i>	
	<i>n</i>	Norm.
<i>give</i>	1	0.56
<i>happen</i>	1	0.56
<i>mean</i>	5	2.78
Total (<i>N</i>)	7	3.89

Table D2a. Verbs Attested in It's + Vø in ACE

<i>Vø</i>	<i>it's + Vø</i>	
	<i>n</i>	Norm.
<i>belong</i>	3	3.00
<i>come</i>	1	1.00
<i>depend</i>	7	7.00
<i>guarantee</i>	1	1.00
<i>have</i>	1	1.00
<i>help</i>	1	1.00
<i>keep</i>	1	1.00
<i>look</i>	1	1.00
<i>make</i>	1	1.00
<i>mean</i>	10	10.00
<i>take</i>	1	1.00
<i>want</i>	1	1.00
Total (<i>N</i>)	29	29.00

Table D2b. Verbs Attested in That's + Vø in ACE

<i>Vø</i>	<i>that's + Vø</i>	
	<i>n</i>	Norm.
<i>give</i>	1	1.00
<i>go</i>	2	2.00
<i>improve</i>	1	1.00
<i>jeopardise</i>	1	1.00
<i>keep</i>	1	1.00
<i>make</i>	1	1.00
<i>mean</i>	3	3.00
<i>provide</i>	1	1.00
<i>take</i>	1	1.00
Total (<i>N</i>)	12	12.00

Table D3a. Verbs Attested in *It's + Vø* in ICNALE

<i>Vø</i>	<i>it's + Vø</i>	
	<i>n</i>	Norm.
<i>affect</i>	4	1.90
<i>bother</i>	1	0.48
<i>bring</i>	1	0.48
<i>cause</i>	1	0.48
<i>change</i>	1	0.48
<i>come</i>	2	0.95
<i>comfort</i>	1	0.48
<i>create</i>	1	0.48
<i>damage</i>	4	1.90
<i>depend</i>	11	5.24
<i>destroy</i>	3	1.43
<i>disturb</i>	4	1.90
<i>do</i>	2	0.95
<i>feel</i>	2	0.95
<i>give</i>	5	2.38
<i>harm</i>	3	1.43
<i>have</i>	5	2.38
<i>help</i>	4	1.90
<i>influence</i>	4	1.90
<i>make</i>	16	7.62
<i>mean</i>	2	0.95
<i>need</i>	3	1.43
<i>prepare</i>	1	0.48
<i>take</i>	2	0.95
<i>taste</i>	2	0.95
<i>warm</i>	1	0.48
Total (<i>N</i>)	86	40.95

Table D3b. Verbs Attested in *That's + Vø* in ICNALE

<i>Vø</i>	<i>that's + Vø</i>	
	<i>n</i>	Norm.
<i>affect</i>	1	0.48
<i>go</i>	2	0.95
<i>have</i>	1	0.48
<i>help</i>	1	0.48
<i>make</i>	4	1.90
Total (<i>N</i>)	9	4.29

Appendix E

Table E1. PMI Score for Verbs Attested in *It's + Vø* and *That's + Vø* in ICE

<i>Vø</i>	PMI Score	Fisher's Exact <i>p</i> -value
<i>coincide</i>	+1.59	0.438
<i>continue</i>	+1.59	0.438
<i>happen</i>	-2.87	0.438
<i>look</i>	+1.59	0.041
<i>mean</i>	-2.32	0.438
<i>remain</i>	+1.59	0.438
<i>take</i>	+1.59	0.438

Table E2. PMI Score for Verbs Attested in *It's + Vø* and *That's + Vø* in ACE

<i>Vø</i>	PMI Score	Fisher's Exact <i>p</i> -value
<i>belong</i>	+1.60	0.543
<i>depend</i>	+2.70	0.085
<i>give</i>	-7.65	0.293
<i>go</i>	-8.64	0.081
<i>improve</i>	-7.65	0.293
<i>jeopardise</i>	-7.65	0.293
<i>keep</i>	-1.27	0.505
<i>make</i>	-1.27	0.505
<i>provide</i>	-7.65	0.293
<i>take</i>	-1.27	0.505

Table E3. PMI Score for Verbs Attested in *It's + Vø* and *That's + Vø* in ICNALE

<i>Vø</i>	PMI Score	Fisher's Exact <i>p</i> -value
<i>affect</i>	-1.26	0.330
<i>bother</i>	-1.54	0.905
<i>bring</i>	-1.54	0.905
<i>cause</i>	-1.54	0.905
<i>change</i>	-1.54	0.905
<i>comfort</i>	-1.54	0.905
<i>create</i>	-1.54	0.905
<i>depend</i>	+1.40	0.313
<i>go</i>	-11.72	0.008
<i>help</i>	-1.26	0.330
<i>make</i>	-1.26	0.071
<i>prepare</i>	-1.54	0.905
<i>warm</i>	-1.54	0.905