

Cross-linguistic influence in upper-intermediate to advanced L2 construction learning: A multidimensional analysis of L1 effects on probability, frequency, diversity, and verb-construction associations

Fatih Ünal Bozdağ and Junhua Mo

*Accepted manuscript. The version of record is published in Lingua, 341, 104239 (2026),
<https://doi.org/10.1016/j.lingua.2026.104239>. Please cite the published version.*

© 2026. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Abstract

Usage-based approaches to second language acquisition treat argument structure constructions (ASCs) as form-meaning pairings whose acquisition depends on learners' repeated exposure to and production of verb-construction combinations. Yet most learner corpus studies of ASCs have examined single dimensions of constructional knowledge in isolation, using mono-factorial approaches that do not fully capture the multiple interacting factors shaping learner production. This study attempts to introduce a multidimensional framework for analyzing argument structure construction knowledge in learner English. Utilizing the International Corpus of Learner English (ICLE), which includes data from 34 first language (L1) backgrounds, we employ Bayesian hierarchical models to examine three dimensions of constructional competence: probability, frequency, and lexical diversity. The results demonstrate varying sensitivities to cross-linguistic influence across these dimensions. L1 effects were strongest for frequency, with 25 out of 34 L1 groups showing credible effects, followed by diversity (18 out of 34) and probability (16 out of 34). Collostructional analysis validates the four construction families (i.e., Motion, Resultative, Transfer and Transitive_Voice) through semantically coherent verb-construction associations. Cross-dimensional dissociations reveal distinct production strategies: German learners exhibit high probability and diversity but low frequency, while Japanese and Turkish learners show elevated frequency but reduced diversity. These findings indicate that deployment frequency is most sensitive to L1 influence and that single-index approaches risk obscuring meaningful variation operating in different directions across dimensions.

Keywords: Learner corpus; argument structure constructions (ASCs); multidimensional analysis; Bayesian hierarchical modeling; cross-linguistic influence; L1 effects

1. INTRODUCTION

Construction Grammar is a usage-based family of linguistic theories that conceptualizes grammatical knowledge as an inventory of form-meaning pairings, or constructions, ranging from morphemes to complex clausal patterns (Goldberg, 1995, 2006; Langacker, 1987). Within this framework, argument structure constructions (ASCs) represent schematic templates specifying how verbs combine with arguments to encode event structures, such as the ditransitive *She gave him a book*, the caused-motion *She pushed the ball into the goal*, or the resultative *He painted the wall red*. For second language (L2) learners, acquiring these constructions involves more than learning abstract schemas. Constructional competence is inherently multidimensional, requiring learners to know that a construction exists in the target language (probability), when to deploy it relative to alternatives (frequency), and which lexical items appropriately populate its slots (diversity). These dimensions need not develop in lockstep and may show differential sensitivity to input properties, instructional context, and first language (L1) influence. Throughout this paper, we use cross-linguistic influence (CLI) and L1 effects interchangeably to refer to the ways in

which learners' first-language grammatical knowledge shapes their production and processing of L2 argument structure constructions.

We propose that entrenchment can be operationalised along three partially dissociable dimensions: strength, threshold, and breadth. This synthesis integrates Bybee's (2010) token-frequency account of entrenchment strength, Langacker's (1987) activation-threshold notion, and Schmid's (2017, 2020) multi-association approach to representational breadth. Although no single source articulates exactly this triad, each component has independent support in the usage-based literature (Diessel, 2019; Divjak, 2019; Ellis, 2002). If L1 constructions transfer these aspects independently, learners from different L1 backgrounds may exhibit qualitatively distinct entrenchment profiles rather than simply "more" or "less" entrenched representations.

Despite growing interest in usage-based approaches to second language acquisition (Bybee, 2010; Ellis, 2012; Tomasello, 2003), previous studies have rarely examined frequency, diversity, and probability as dimensions of constructional knowledge simultaneously within a unified framework that accounts for their potential dissociation. Frequency-based analyses (e.g., Gries, 2018; Kyle, 2016) reveal how often learners produce particular constructions but overlook whether production draws on narrow or broad lexical repertoires. Diversity-based approaches, rooted in complexity-accuracy-fluency (CAF) research (Housen et al., 2012; Norris & Ortega, 2009), capture lexical range but may conflate high-diversity production with mere high-frequency output. As Gries (2015, 2021) has argued, aggregate corpus analyses that collapse data across learners and texts can obscure individual variation and strain the independence assumptions underlying standard statistical tests. Treating nested observations as independent underestimates standard errors and inflates Type I error rates, meaning that prior reports of L1 background effects may overstate the strength of cross-linguistic differences (Barr et al., 2013; Cunnings, 2012). Bayesian hierarchical models address this directly by estimating learner- and text-level variance components, producing uncertainty intervals that reflect the true structure of the data. These limitations motivate a multidimensional framework capable of disentangling complementary facets of constructional knowledge while accommodating the hierarchical structure of learner corpus data.

The present study develops a framework that operationalizes three dimensions of constructional knowledge using Bayesian hierarchical models: (1) probability, which captures whether learners produce a construction, (2) frequency, which reflects the intensity of deployment conditional on use, and (3) lexical diversity, which measures the breadth of verb repertoires within a construction family. Bayesian hierarchical models jointly estimate group-level and individual-level effects while propagating uncertainty throughout the inference (Norouzian et al., 2018); unlike null-hypothesis testing, they produce full posterior distributions for each parameter, enabling direct probabilistic statements about effect sizes. By separating these dimensions analytically, this study aims to reconcile conflicting findings in prior literature regarding the persistence of L1 effects at advanced proficiency levels and to provide diagnostic indices that separate construction availability from deployment preference and lexical breadth. The framework has potential applications in L2 assessment and pedagogy: learners with similar overall proficiency scores may differ along the three dimensions in ways that single-index measures cannot detect, and instructional interventions can therefore be targeted to the specific dimension (access, frequency or diversity) in which a given learner or L1 group shows a gap. We return to these implications in Section 6.2.3.

2. LITERATURE REVIEW

2.1. Developmental trajectories in ASC acquisition

Research using learner corpus data reveals consistent developmental patterns in L2 learners' production of argument structure constructions, indicating a shift from concrete, item-based knowledge to abstract, productive schemas. At lower proficiency levels, learners tend to use a limited set of high-frequency constructions with prototypical verbs (Ellis & Ferreira-Junior, 2009; Eskildsen, 2012; Kyle, 2016). Ellis and Ferreira-Junior's (2009) longitudinal data show that learners' initial ASC repertoires mirror the Zipfian input distributions they are exposed to, with pathbreaking verbs (e.g., *give*, *put* and *go*) disproportionately anchoring the ditransitive, caused-motion, and intransitive-motion constructions during early acquisition. Their findings align with Goldberg et al.'s (2004) experimental demonstration that skewed input distributions support the formation of constructional categories. Evidence for this facilitative effect is not uniform, however: Year and Gordon (2009) found that Korean L2 English learners acquired the ditransitive more effectively under balanced input than under input skewed toward the prototypical verb *give*, suggesting that pathbreaking-verb effects may be moderated by L1 typological profile. This conservatism highlights item-based learning, where constructions are acquired as specific chunks linked to particular verbs.

As proficiency increases, learners use a broader range of argument structure constructions and demonstrate greater lexical diversity, moving beyond prototypical verbs to include more varied, lower-frequency options (Hwang & Kim, 2022; Liu & Lu, 2024). Römer and Berger (2019) observed that German and Spanish learners showed systematic increases in constructional diversity, productivity, and native-likeness across proficiency levels, supporting universal usage-based learning mechanisms. Higher-proficiency learners effectively extend constructions to non-prototypical verbs, indicating the acquisition of abstract semantic templates (e.g., *X causes Y to become Z*) as productive grammatical resources, reflecting a shift from item-based to schema-based knowledge.

While these developmental trajectories suggest universal usage-based patterns, the rate and nature of constructional development are strongly modulated by learners' first language, which raises the question of how L1 typology shapes L2 construction acquisition.

2.2. Cross-linguistic influence on constructional choice

L2 construction acquisition is influenced by learners' first language. Cross-linguistic influence is a significant finding in corpus-based construction research. Talmy's (2000) typological distinction between satellite-framed languages (e.g., English and German) and verb-framed languages (e.g., Spanish and French) shapes L2 acquisition patterns (Ibarretxe-Antuñano, 2009; Slobin, 1996, 2004). In a prototypical satellite-framed language like English, manner of motion is encoded in the verb while path is carried by a satellite, as in *The bottle floated into the cave*, whereas a verb-framed language like Spanish encodes path in the verb and manner in a separate constituent, as in *La botella entró a la cueva flotando* ('*The bottle entered the cave floating*'). Learners whose L1 is verb-framed therefore tend to default to general path verbs and leave manner implicit, while satellite-framed learners may underexploit path satellites when producing verb-framed target structures (Cadierno, 2008; Cadierno & Ruiz, 2006; Jarvis & Pavlenko, 2008). Learners from satellite-framed backgrounds acquire caused-motion and intransitive motion constructions more easily, using them more often (Cadierno, 2008; Cadierno & Ruiz, 2006; Römer et al., 2014).

Typological transfer mechanisms involve attentional and processing biases shaped by L1 experience (Ellis, 2006; Slobin, 1996). Ellis (2006) suggested that L1 constructions create

persistent *learned attention* patterns during L2 acquisition, causing learners to focus on reliable cues from their L1 and potentially overlook different L2 cues. Verb-framed L1 speakers may prioritize path information in verbs, neglecting manner details in satellites, which hinders their use of English manner verbs in caused-motion frames. L2 learners more generally show greater processing cost when verbs are weakly associated with a given argument structure construction, with longer reading times reflecting less established form-meaning mappings (Kim & Shin, 2025).

Cross-linguistic influence also occurs at the micro-level through specific lexical-constructional associations, known as *collostructional transfer* (Goschler & Stefanowitsch, 2024; Gries & Wulff, 2005). The statistical regularities of L1 can influence L2 production, resulting in grammatically correct but idiomatically awkward utterances. Using acceptability judgment tasks with Turkish learners of English, Gedik and Uslu (2023) showed that judgments of English ditransitive combinations are inflated for verb-construction pairs strongly attracted in the Turkish ditransitive, even when those same combinations are weakly attracted or repelled in English, demonstrating a constructional transfer effect that persists into advanced proficiency. This indicates that deeply entrenched L1 grammar can bias L2 receptive knowledge, particularly for verbs with high L1 attraction to a construction when corresponding L2 pairings are less frequent or weakly established.

Cross-linguistic influence is gradient and evolves with increased L2 proficiency and exposure. As learners improve, their verb-construction associations align more closely with native patterns (Römer & Berger, 2019). The timeline for overcoming transfer varies based on construction salience and input frequency. High-frequency constructions with explicit pedagogical focus exhibit minimal lasting transfer, while low-salience constructions may retain L1 influences even at advanced levels. Additionally, the aspect of constructional knowledge examined can affect transfer detection; studies on repertoire may show minimal L1 effects, while those focusing on lexical choices may reveal persistent cross-linguistic patterns.

These cross-linguistic mechanisms, however, do not apply uniformly across construction types: different argument structure constructions present different acquisition profiles, and the next section surveys the construction-specific findings that motivate our four-family taxonomy.

2.3. Construction-specific patterns in learner production

General developmental principles and cross-linguistic influences vary across different argument structure constructions, notably in the dative alternation, which involves choosing between double-object (e.g., *She gave him a book*) and prepositional dative constructions (e.g., *She gave a book to him*). L2 learners from diverse L1 backgrounds generally prefer the double-object construction, aligning with native speaker tendencies, while developing sensitivity to probabilistic constraints as proficiency increases (Callies & Szczesniak, 2008). However, L1 transfer remains significant; for example, Arabic learners tend to favor the prepositional dative due to grammatical constraints in Arabic against double-object structures (Aljadani, 2018). Bozdağ (2024) found that L1 background had minimal explanatory power beyond general inter-learner variation, suggesting that universal processing constraints may dominate at the alternation level, but the impact of transfer on verb selection remains uncertain.

Caused-motion and resultative constructions are notably influenced by L1 typology, particularly affecting learners from verb-framed backgrounds. For instance, Korean, Spanish, and French learners underuse these constructions compared to native speakers and those from satellite-framed backgrounds (Cadierno, 2008; Römer et al., 2014). Verb selection patterns further illustrate this influence, as verb-framed learners often choose general-purpose motion verbs (e.g., *put* and

go) instead of specific manner-of-motion verbs (e.g., *shove* and *slide*) common in native English. Resultative constructions are especially challenging for L2 learners due to their structural complexity and low frequency, and their acquisition involves extending schema-based knowledge from prototypical pairings (e.g., *paint something red*) to novel verb-resultant combinations (Boyd & Goldberg, 2009; Goldberg, 1995).

These construction-specific findings underscore the need for analytical frameworks capable of capturing multiple dimensions of constructional knowledge simultaneously, rather than collapsing learner production into a single index.

2.4. Towards multidimensional assessment of L2 constructional knowledge

Despite significant empirical progress, methodological limitations hinder a comprehensive understanding of L2 constructional knowledge. Gries (2018) noted that corpus-wide aggregate analyses are inadequate for complex observational data, as they treat all observations as independent and overlook dispersion effects and individual variation. Additionally, monofactorial methods isolate single-factor influences without accounting for confounding variables, a concern also relevant to standard constructional analysis. In response, learner construction research is increasingly utilizing mixed-effects and Bayesian regression models, which accommodate hierarchical data structures, allow for simultaneous assessment of multiple predictors, and quantify uncertainty in parameter estimates (Cunnings, 2012; Gries, 2015, 2021).

A significant limitation in current research is the unidimensional operationalization of constructional knowledge, where studies typically examine a single index such as frequency counts or type-token ratios in isolation (Housen & Kuiken, 2009; Housen et al., 2012; Norris & Ortega, 2009; Verspoor et al., 2012). More general multi-factor accounts of SLA have also argued for integrating multiple interacting principles rather than privileging any single factor (Filipović & Hawkins, 2013). However, usage-based theory suggests that construction knowledge is multifaceted, involving distinct competencies: awareness of a construction's existence (probability), knowing when to use it (frequency), and recognising which lexical items fit (diversity). The complexity-accuracy-fluency (CAF) framework in L2 writing research similarly demonstrates that linguistic development is multidimensional and progresses at different rates along separable dimensions. Our three dimensions map onto the triad of strength, threshold, and breadth introduced in Section 1, synthesising token-frequency, activation-threshold, and associative-breadth accounts in the usage-based literature (Bybee, 2010; Diessel, 2019; Divjak, 2019; Langacker, 1987; Schmid, 2017, 2020). Single-index approaches risk conflating these competencies, potentially leading to contradictory findings.

3. METHODS

3.1. Research questions

The study addresses how L1 influence operates across three dimensions of construction knowledge in upper-intermediate to advanced L2 learners, using Bayesian hierarchical models to estimate probability, frequency, and lexical diversity effects jointly while accounting for the nested structure of learner corpus data. It poses three research questions.

RQ1: How are probability, frequency, and lexical diversity distributed across argument structure construction families in upper-intermediate to advanced learner English corpora?

RQ2: To what extent do these three dimensions exhibit independent variation across L1 backgrounds, and at which construction families do the dimensions most clearly dissociate?

RQ3: Which L1 backgrounds show credible effects on each dimension, and what does the cross-dimensional pattern reveal about the locus of cross-linguistic influence in upper-intermediate to advanced L2 production?

3.2. *Learner corpus and L1 participants*

This study analyzed the International Corpus of Learner English (ICLE) V3 (Granger et al., 2020), a widely-used corpus of argumentative essays written by upper-intermediate to advanced L2 English learners (approximate CEFR B2- C1). The corpus comprises 9,529 essays from 34 L1 backgrounds, totaling approximately 5.7 million words. All participants were university-level learners aged 18–25, with formal instruction in English ranging from 8–15 years. Essays were written under timed conditions (60–90 minutes) on topics related to social issues, education, and technology. Because all L1 groups wrote on the same standardised topic set, topic selection is controlled across L1s; nevertheless, the argumentative genre systematically favours transitive-voice constructions over manner-of-motion or caused-motion constructions, a property that may partly explain the dominance of transitive-voice constructions in our corpus (we return to this issue in Section 6.3).

The 34 L1 backgrounds represented diverse typological profiles, including Indo-European languages (e.g., Germanic: German, Dutch, Norwegian and Swedish; Romance: French, Italian, Spanish and Portuguese; Slavic: Russian, Polish, Czech, Bulgarian, Serbian, Bosnian and Macedonian; Indo-Aryan: Urdu and Punjabi), Uralic languages (e.g., Finnish and Hungarian), Sino-Tibetan languages (e.g., Chinese-Mandarin and Chinese-Cantonese), Turkic (i.e., Turkish), Japonic (i.e., Japanese), and Koreanic (i.e., Korean), Afro-Asiatic languages (e.g., Arabic), Niger-Congo languages (e.g., Tswana), and others (e.g., Greek, Albanian, Lithuanian and Persian). This typological diversity enabled investigation of cross-linguistic influence patterns across major language families (see Table 1).

Table 1
L1 backgrounds in ICLE V3 grouped by language family and motion-event typology.

Language family	L1 backgrounds in ICLE V3	N L1s
Germanic (S-framed)	German, Dutch, Norwegian, Swedish	4
Romance (V-framed)	French, Italian, Spanish, Portuguese	4
Slavic (S-framed)	Russian, Polish, Czech, Bulgarian, Serbian, Bosnian, Macedonian	7
Indo-Aryan (S-framed)	Urdu, Punjabi	2
Uralic	Finnish, Hungarian	2
Sino-Tibetan (equipollently-framed)	Chinese-Mandarin, Chinese-Cantonese	2
Turkic (V-framed)	Turkish	1
Japonic (V-framed)	Japanese	1
Koreanic (V-framed)	Korean	1
Afro-Asiatic	Arabic	1
Niger-Congo	Tswana	1
Hellenic	Greek	1
Albanian	Albanian	1
Baltic	Lithuanian	1
Iranian	Persian	1
15 language families in total	34 L1s in total	

Note: (1) Typological profiles follow Talmy (2000); (2) Per-L1 essay counts and word counts for each subcorpus are available from the ICLE V3 documentation (Granger et al., 2020) and are reported in full in Supplementary Table S5.

3.3. Data retrieval

We developed a three-stage annotation pipeline to identify and extract argument structure constructions from the ICLE. The pipeline leveraged a state-of-the-art pre-trained construction tagger combined with post-processing for lemmatization and family-level aggregation.

Stage 1: ASC Tagger construction extraction. All essays were processed using the ASC Tagger V2 (Sung & Kyle, 2024), a transformer-based spaCy model fine-tuned on manually annotated construction data. The ASC Tagger automatically identifies nine fine-grained argument structure construction types using Named Entity Recognition (NER) architecture: TRAN_S (Transitive Subject), PASSIVE (Passive Voice), DITRAN (Ditransitive), INTRAN_MOT (Intransitive Motion), CAUS_MOT (Caused Motion), INTRAN_RES (Intransitive Resultative), TRAN_RES (Transitive Resultative), ATTR (Attributive/Copular), and INTRAN_S (Simple Intransitive). The model achieves state-of-the-art performance on construction identification tasks and has been validated on multiple learner corpora (Sung & Kyle, 2024).

Stage 2: Construction selection and lemmatization. From the nine construction types identified by the ASC Tagger, we excluded two categories that were not theoretically relevant to our research questions: ATTR (Attributive/Copular constructions such as *She is a teacher*) and INTRAN_S (Simple Intransitive constructions without directional or resultative semantics, such as *He laughed*). Simple intransitive constructions in particular carry minimal event-structural information and show limited typological variation in argument realisation, unlike the motion, resultative, transfer, and transitive constructions that systematically participate in the satellite-versus verb-framed distinction central to our research questions. These exclusions reduced noise from high-frequency copular verbs (e.g., *be*, *is* and *are*) and semantically underspecified intransitives, allowing us to focus on argument structure constructions with richer event semantics and cross-linguistic variation potential.

For the seven retained construction types, we extracted clean verb lemmas by re-parsing each essay with spaCy v3.7.2 using the `en_core_web_trf` model (Honnibal et al., 2020). We used the character offsets provided by the ASC Tagger to locate the corresponding verb token in the spaCy parse tree, and then retrieved its lemma form and part-of-speech tag. This offset-based approach prevents token contamination issues that can arise from misaligned entity boundaries. This two-model approach combines the ASC Tagger’s specialized construction knowledge with spaCy’s robust lemmatization. We validated lemmatization quality by manually checking 1,000 randomly sampled verb instances. Precision reached 98.30% (983 correct out of 1,000), with 17 errors stemming primarily from misspellings in learner text. Recall was not separately estimated because the validation sample was drawn from the positive set rather than a balanced gold standard.

To distinguish main verbs from auxiliary verbs (a critical distinction for argument structure analysis), we relied on spaCy’s part-of-speech tagger, which Honnibal et al. (2020) report at approximately 92% F1 on VERB vs. AUX classification against the OntoNotes benchmark. We retained only tokens tagged as VERB, filtering out auxiliary uses of *have*, *do*, and *be*. Across the full dataset, auxiliary tokens represented 8.00% of all verbs, with *have* (possessive) and *do* (support) contributing the majority of contamination cases. Because contamination is concentrated in a predictable verb subset and affects all L1 groups equally, it introduces noise but no systematic bias. This level of concentrated contamination is comparable to precision rates reported for other

automated construction annotation pipelines: Sung and Kyle (2024), the ASC Tagger developers, report weighted F1 scores of 0.86–0.93 across construction types in their validation study, and Kyle (2016) similarly argues that systematic filtering approaches yield data of sufficient quality for publication when errors are concentrated in predictable classes and affect all groups equally. For descriptive statistics of lexical diversity (verb types, type-token ratios and Shannon entropy), a minimum frequency threshold of 10 corpus-wide occurrences was applied to exclude hapax legomena and spelling errors in learner text.

Stage 3: Family-level aggregation. The seven fine-grained construction types were aggregated into four coarse families based on shared semantic and syntactic properties, following theoretical distinctions in argument structure research (Goldberg, 2006; Levin, 1993):

Motion: INTRAN_MOT (e.g., *The ball rolled into the goal*); CAUS_MOT (e.g., *She pushed the ball into the goal*)

Resultative: INTRAN_RES (e.g., *The door swung open*); TRAN_RES (e.g., *He painted the wall red*)

Transfer: DITRAN (e.g., *She gave him a book*)

Transitive_Voice: TRAN_S (e.g., *She wrote a letter*); PASSIVE (e.g., *The letter was written*)

The Transitive_Voice aggregation warrants specific justification. Active transitive and passive realizations draw on substantially overlapping verb inventories and represent voice alternations of the same underlying argument structure (Agent-Theme). Because active and passive realizations draw from a common pool of passivizable transitive verbs, separating them would create largely redundant collostructional profiles and artificial sparsity in diversity estimates, especially for the much rarer passive realizations. This aggregation aligns with our focus on lexical implementation of argument structure rather than syntactic voice choice. However, this design decision means that voice-specific L1 effects cannot be detected; we return to this limitation in Section 6.3. An empirical assessment of verb overlap supports this aggregation: among the unique verbs appearing in TRAN_S and PASSIVE, the two categories draw from a substantially shared lexical pool, yielding a Jaccard similarity of 0.32. Combined with the substantial size asymmetry between the two realizations, this overlap indicates that a disaggregated analysis would treat voice alternations of the same underlying argument structure as statistically independent, inflating the apparent lexical profile of the rarer passive construction.

The ASC Tagger initially identified 847,859 construction spans across all nine construction types. After excluding ATTR (Attributive/Copular) and INTRAN_S (Simple Intransitive) categories as described above, the final dataset comprised 560,240 construction spans across the four target families, with each span annotated for: (a) construction family (i.e., Motion, Resultative, Transfer and Transitive_Voice), (b) verb lemma (e.g., give, make and bring), (c) essay ID, and (d) L1 background.

To illustrate what the data look like after ASC tagging and lemmatization, Table 2 shows six representative pipeline outputs, one per construction type retained for analysis, using the example sentences introduced in Stage 3 above. Each row reports the source excerpt, the ASC Tagger label, the verb token and its spaCy lemma, and the construction family to which the type is aggregated.

Table 2

Illustrative pipeline output after ASC tagging and lemmatization, using the schema sentences introduced in Stage 3.

Source excerpt	ASC label	Verb token	Lemma	Family
----------------	-----------	------------	-------	--------

The ball rolled into the goal.	INTRAN_MOT	rolled	roll	Motion
She pushed the ball into the goal.	CAUS_MOT	pushed	push	Motion
He painted the wall red.	TRAN_RES	painted	paint	Resultative
She gave him a book.	DITRAN	gave	give	Transfer
She wrote a letter.	TRAN_S	wrote	write	Transitive_Voice
The letter was written.	PASSIVE	written	write	Transitive_Voice

Note: (1) Verb token is the surface form identified by the ASC Tagger; (2) Lemma is the spaCy v3.7.2 en_core_web_trf lemma form.

Table 3 summarises the yields at each pipeline stage.

Table 3
Data at each pipeline stage.

Stage	Data	N Spans	N Essays	N L1s
0	ICLE V3 raw essays	—	9,529	34
1	ASC-tagged spans (9 construction types)	847,859	9,529	34
2	ASC-tagged spans after excluding ATTR and INTRAN_S (7 types)	560,240	9,529	34
3	Family aggregation (4 families: Motion, Resultative, Transfer, and Transitive_Voice)	560,240	9,529	34

3.4. Data analysis

3.4.1. Descriptive statistics of construction families

We computed descriptive statistics to characterize the distribution of construction families across the corpus. For each family $f \in \{\text{Motion, Resultative, Transfer, and Transitive_Voice}\}$, we calculated:

Absolute frequency: the total count of construction tokens in the corpus

Relative frequency: the proportion of all construction tokens belonging to family f

Coverage: the percentage of all essays containing at least one instance of family f

Verb type count: the number of unique verb lemmas attested in family f

These statistics offered a corpus-level overview of construction usage patterns and informed subsequent modeling decisions, such as identifying which families were sufficiently frequent for reliable statistical inference.

3.4.2. Collostructional analysis

To quantify verb-construction associations, we conducted collostructional analysis (Stefanowitsch & Gries, 2003), a corpus-linguistic method that measures the statistical attraction or repulsion between lexical items (verbs) and constructions. For each verb-family pair, we constructed a 2×2 contingency table and applied Fisher’s exact test, yielding an exact p -value for the null hypothesis of independence. We report the negative $\log_{10}$ -transformed p -value ($-\log_{10} p$), where larger values indicate stronger associations.

In addition to p -values, we computed $\Delta P(F|V)$, defined as the difference between $P(\text{family} | \text{verb})$ and $P(\text{family})$, which measures how much observing a given verb increases the probability

of the construction family. Results were ranked by $-\log_{10} p$ to identify the top collostructions per family.

3.4.3. Bayesian hierarchical regression

To assess L1 influence on construction usage while accounting for the hierarchical structure of corpus data, we developed three Bayesian hierarchical regression models targeting complementary facets of construction knowledge. All models were estimated using PyMC v5.10.4 (Abril-Pla et al., 2023) with the No-U-Turn Sampler (Hoffman & Gelman, 2014). For all three models, we ran 4 independent MCMC chains with 2,000 tuning steps and 2,000 sampling steps per chain, yielding 8,000 posterior samples. All parameters satisfied convergence criteria: $\hat{R} \leq 1.01$, $ESS > 1,000$, and zero divergent transitions. Following standard Bayesian practice (Kruschke, 2018; McElreath, 2020), we considered an L1 effect credible when its 94% Highest Density Interval (HDI) excluded zero. The HDI represents the range of parameter values with the highest posterior probability, and exclusion of zero indicates that we can be 94% confident the effect is meaningfully different from no effect. Full model specifications, including likelihood functions, link functions, and prior distributions, are provided in Supplementary Materials S1¹.

3.4.3.1. Model 1: Probability

Model 1 estimated the probability that an essay contains at least one token of a given construction family. The data structure consisted of 38,116 essay-family observations (9,529 essays $\times$ 4 families), with binary outcomes indicating presence (83.59%) or absence (16.41%) of each family. We specified a Bernoulli model with complementary log-log link, which is theoretically appropriate for probability data with ceiling effects (some families appear in $> 90.00\%$ of essays). L1 background entered as fixed effects (34 levels, sum-to-zero contrast), family as random effects (4 levels), with centered log word count as an offset to control for essay length.

3.4.3.2. Model 2: Frequency

Model 2 targeted count outcomes at the essay $\times$ verb $\times$ family level, comprising 375,647 observations. We specified a Poisson model with log link, where L1 background and family entered as fixed effects, verb entered as random effects to capture lexical idiosyncrasy, and log word count entered as an offset so that coefficients represent effects on construction density (tokens per word) rather than raw counts. Posterior predictive checks confirmed that Poisson assumptions were appropriate (overdispersion parameter near zero).

3.4.3.3. Model 3: Diversity

Model 3 targeted lexical diversity, operationalised as Shannon entropy (bits) per essay per family. Because diversity scores exhibited heavy tails, we specified a Student- t regression with estimated degrees-of-freedom parameter, providing robustness to outliers. L1 background entered as fixed effects, family as random effects, with log token count as a covariate controlling for opportunity effects (essays with more construction tokens have more opportunities for verb variety).

3.4.4. Sensitivity analysis

¹ All the supplementary materials are publicly available at https://osf.io/bg39s/overview?view_only=7212dbc06b99462898527173691d3e0e.

To assess whether L1 patterns were influenced by the dominant Transitive_Voice family, we re-ran all three models excluding Transitive_Voice and retaining only Motion, Resultative, and Transfer constructions. The model specifications remained identical to those in the primary analyses (Sections 3.4.1–3.4.3), with family random effects (M1 and M3) or fixed effects (M2) adjusted for three levels instead of four. This adjustment reduced the dataset from 375,647 observations to 61,120 observations, resulting in an 83.73% reduction for M2. We compared the credibility of L1 effects between the original and sensitivity models by counting how many L1 groups showed 94% HDIs excluding zero in each analysis. Full results can be found in Supplementary Materials S2.

4. RESULTS

Before presenting the results, we briefly recall the three research questions that motivate the analyses and indicate where each is addressed. The first (RQ1) concerns how probability, frequency, and lexical diversity are distributed across the argument structure construction families, and is taken up through the descriptive statistics in Section 4.1. The second (RQ2) asks to what extent these dimensions vary independently across L1 backgrounds and where they most clearly dissociate, examined via the verb-construction associations in Section 4.2. The third (RQ3) concerns which L1 backgrounds show credible effects on each dimension and what the cross-dimensional pattern reveals about the locus of cross-linguistic influence, addressed through the Bayesian hierarchical models in Section 4.3.

4.1. Overall use of ASCs

This section reports descriptive statistics for the distribution of construction families across the ICLE corpus, followed by collostructional analysis of verb-construction associations in Section 4.2 and Bayesian hierarchical models of L1 effects in Section 4.3.

As shown in Table 4, there are a total of 560,240 annotated construction tokens across the four argument-structure families, with a strongly skewed distribution that favors the Transitive family (including both active and passive realizations of Transitive_Voice). Specifically, Transitive accounts for 478,795 tokens (approximately 85.46% of all construction tokens), while Resultative contributes 37,767 (6.74%), Motion 34,493 (6.16%), and Transfer 9,185 (1.64%).

Table 4
Tokens and percentage of construction family.

	Motion	Resultative	Transfer	Transitive	Total
Tokens	34,493	37,767	9,185	478,795	560,240
Percentage (%)	6.16	6.74	1.64	85.46	100.00

As shown in Table 5, the skewed distribution is also evident in coverage: Transitive_Voice appears in every essay (100.00% coverage), while Motion and Resultative appear in the vast majority of essays (92.33% and 91.91%, respectively). In contrast, Transfer is present in only about half of the essays (50.14%), resulting in a high zero-usage share of 49.86%.

Table 5
Corpus scale and coverage by construction family.

Family	Tokens	Coverage (%)	Verb types	TTR
Motion	34,493	92.33	808	.023

Resultative	37,767	91.91	448	.012
Transfer	9,185	50.14	169	.018
Transitive_Voice	478,795	100.00	1,696	.004

Note: (1) Coverage = percentage of essays containing ≥ 1 token; (2) Verb Types = unique verb lemmas; (3) TTR = type-token ratio; (4) Verb type counts and TTR are computed after excluding verbs with fewer than 10 corpus-wide occurrences to remove misspellings and noise from learner text (548,036 tokens retained, 97.80% of total).

As shown in Table 6, median per-essay rates per 1,000 words are 84.05 for Transitive_Voice, with a central 80.00% band roughly spanning 65.01 to 105.07. In comparison, Resultative centers around 6.38 per 1,000 words (≈ 2.18 to 13.06), Motion around 5.69 (≈ 2.04 to 11.91), and Transfer around 2.29 (≈ 1.36 to 6.02).

Table 6

Per-essay construction rates per 1,000 words.

Family	Median	Mean	SD	P10	P90
Motion	5.69	6.57	4.23	2.04	11.91
Resultative	6.38	7.20	4.37	2.18	13.06
Transfer	2.29	3.23	2.35	1.36	6.02
Transitive_Voice	84.05	84.67	16.05	65.01	105.07

Note: (1) Statistics computed on essays containing ≥ 1 token per family; (2) P10/P90 = 10th/90th percentiles.

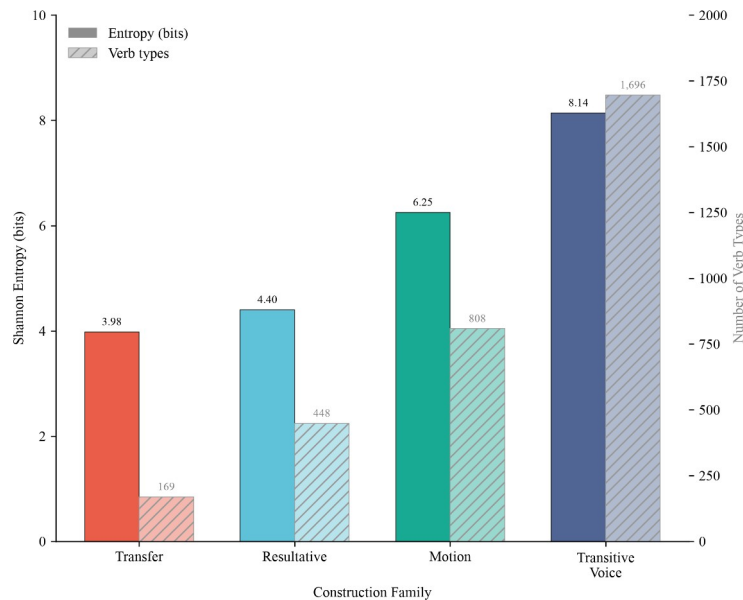


Fig. 1. Lexical diversity by family.

As Figure 1 shows, type-token profiles reveal a broad lexical base for Transitive_Voice and narrower, but still substantial, repertoires for the other families. Measured over the pooled verb lemmas, Transitive_Voice exhibits the highest diversity (entropy ≈ 8.14 bits) and the largest verb

inventory (1,696 lemmas), followed by Motion (6.25 bits; 808 lemmas), Resultative (4.40 bits; 448 lemmas), and Transfer (3.98 bits; 169 lemmas).

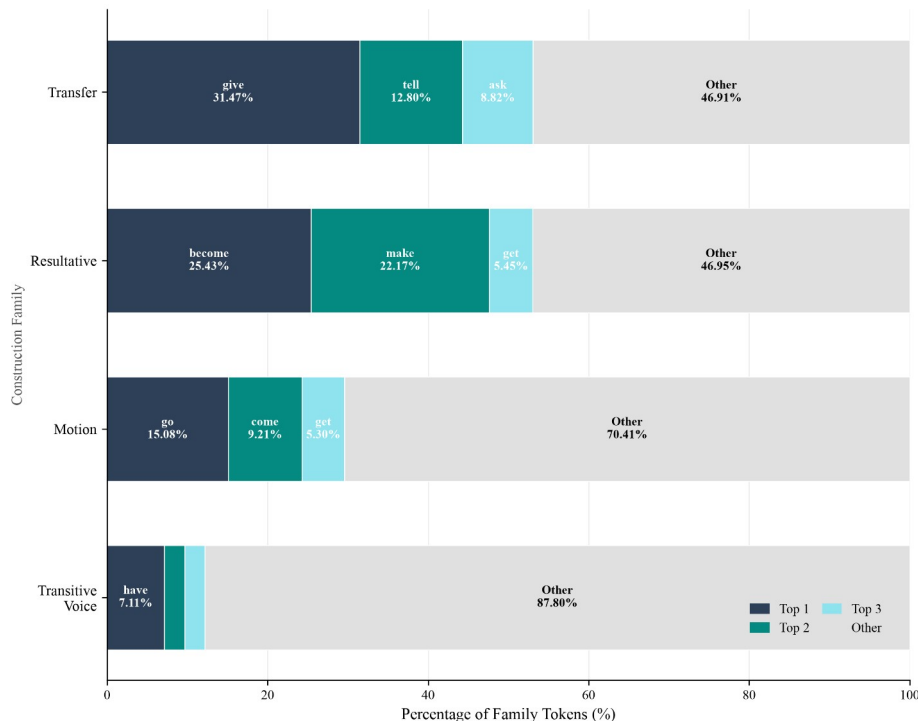


Fig. 2. Percentage of Family Tokens (%).

Note: (1) Percentages computed over verbs appearing in at least 3 texts (553,224 tokens); (2) Percentages computed over verbs with family frequency ≥ 2 ($n = 33,510 / 37,236 / 9,034 / 473,444$).

As Figure 2 shows, the top-verb profiles align with these distributions while highlighting family-specific preferences. Within Motion, the most frequent lemmas are *go* ($\approx 15.08\%$), *come* ($\approx 9.21\%$), and *get* ($\approx 5.30\%$). Resultative is led by *become* ($\approx 25.43\%$), *make* ($\approx 22.17\%$), and *get* ($\approx 5.45\%$). Transfer is headed by *give* ($\approx 31.47\%$), *tell* ($\approx 12.80\%$), and *ask* ($\approx 8.82\%$). For Transitive_Voice, the long tail is headed by *have* ($\approx 7.11\%$), *think* ($\approx 2.58\%$), and *do* ($\approx 2.51\%$).

4.2. Verb–construction associations

We conducted Fisher’s exact collocation analysis (Stefanowitsch & Gries, 2003) over verb–construction pairs across four construction families. Associations are summarized as evidence ($-\log_{10} p$) and $\Delta P(F|V) = P(F|V) - P(F)$, where $P(F)$ is the family’s corpus baseline and $P(F|V)$ is the conditional probability of the family given the verb.

Table 7 presents the top associations for each family. Motion constructions display exceptional semantic coherence. That is, Motion verbs including the path verb *return* and manner verbs such as *travel* and *walk* exhibit high ΔP values (.80–.87), indicating strong specificity to motion frames, while *think* and *use* show repulsion. Resultative constructions attract change-of-

state verbs (e.g., *keep*, *grow* and *force*) while repelling transfer verbs (e.g., *give*). Transfer constructions show a highly specific profile dominated by communication verbs (e.g., *show*, *offer* and *remind*). Transitive_Voice shows cognitive predicates as attracted (e.g., *believe*, *mean* and *need*) and intransitive-preferring verbs as repelled (e.g., *fall*, *remain* and *move*). The latter pattern likely reflects a mixture of construction boundary extensions in learner production and occasional parsing noise.

Table 7
Top verb-construction associations by family.

Family	Verb	$-\log_{10} p$	Log Odds	$\Delta P(F V)$	Direction
Motion	return	317.75	4.56	0.80	Attracted
	travel	287.72	5.38	0.87	Attracted
	escape	268.12	3.32	0.58	Attracted
	walk	245.80	5.01	0.85	Attracted
	think	315.77	-4.19	-0.06	Repelled
	use	302.60	-4.92	-0.06	Repelled
Resultative	keep	278.74	1.68	0.21	Attracted
	grow	188.55	2.98	0.52	Attracted
	force	170.46	1.99	0.28	Attracted
	come	146.55	1.13	0.12	Attracted
	give	204.57	-2.88	-0.06	Repelled
	see	170.17	-2.93	-0.06	Repelled
Transfer	show	292.62	2.23	0.11	Attracted
	offer	214.61	2.66	0.17	Attracted
	remind	142.65	4.12	0.49	Attracted
	convince	83.35	3.61	0.36	Attracted
	make	111.19	-3.80	-0.02	Repelled
	do	45.97	-1.73	-0.01	Repelled
Transitive_Voice	believe	284.32	7.26	0.15	Attracted
	mean	282.32	5.09	0.15	Attracted
	need	277.55	4.88	0.15	Attracted
	learn	270.67	3.63	0.14	Attracted
	fall	314.68	-3.30	-0.68	Repelled
	remain	284.19	-4.53	-0.80	Repelled

Note: (1) $\Delta P(F|V) = P(\text{Family}|\text{Verb}) - P(\text{Family})$; (2) Positive values indicate the verb increases family probability (attraction), while negative values indicate the verb decreases family probability (repulsion); (3) Log odds indicates direction and magnitude: positive = attraction, while negative = repulsion.

Cross-family comparison revealed differential degrees of lexical entrenchment. Motion and Transitive_Voice constructions showed the steepest initial associations ($-\log_{10} p > 300$), indicating a small number of highly distinctive verbs, followed by gradual decline. Transfer constructions, despite lower overall frequency, exhibited a more gradual decay pattern, suggesting more distributed lexical associations. L1-specific collostructional patterns revealed both

typologically predictable and unexpected cross-linguistic variation (see Supplementary Table S1 for complete rankings).

4.3. L1 influence in the Bayesian hierarchical models

We fitted three hierarchical Bayesian models targeting complementary facets of construction use: occurrence probability (M1), token frequency conditional on use (M2), and lexical diversity (M3). All models exhibited excellent convergence (all $\hat{R} \leq 1.01$, no divergences) with effective sample sizes exceeding recommended thresholds (Vehtari et al., 2021) ($ESS > 400$ for all parameters, with most exceeding 1,000). Tables 8 to 10 report, for each model, the L1 groups with the largest-magnitude credible effects in each direction (94% HDI excluding zero); complete posterior summaries for all 34 L1 groups are provided in Supplementary Tables S2 to S4.

4.3.1. Model 1: Probability

Model 1 estimated the probability that an essay contains at least one token of a given construction family, using 38,116 essay-family observations.

As shown in Table 8, there is moderate variation in L1 effects: 16 out of the 34 L1 groups (47.06%) exhibited credible effects on construction probability. The strongest positive effect came from Tswana ($\beta = 0.351$), exceeding all Germanic L1s. Germanic L1s also demonstrated elevated probabilities (German $\beta = 0.234$, Dutch $\beta = 0.228$, Swedish $\beta = 0.196$), while several East Asian, Romance, and Baltic L1s exhibited reduced probabilities (Japanese $\beta = -0.142$, Korean $\beta = -0.134$, Spanish $\beta = -0.145$, Chinese-Cantonese $\beta = -0.182$, Lithuanian $\beta = -0.200$). This partial typological pattern aligns with predictions from motion event typology; however, exceptions — particularly Tswana's positive effect — complicate a straightforward typological account.

Table 8

L1 groups with the largest credible effects on Model 1.

L1 background	β	94% HDI	Effect
Tswana	0.351	[0.26, 0.45]	1.42
German	0.234	[0.14, 0.34]	1.26
Dutch	0.228	[0.10, 0.36]	1.26
Swedish	0.196	[0.09, 0.30]	1.22
Korean	-0.134	[-0.24, -0.03]	0.87
Japanese	-0.142	[-0.24, -0.04]	0.87
Spanish	-0.145	[-0.26, -0.03]	0.87
Chinese-Cantonese	-0.182	[-0.27, -0.09]	0.83
Lithuanian	-0.200	[-0.30, -0.09]	0.82

Note: (1) β = effect on complementary log-log scale; (2) Effect = $\exp(\beta)$ represents the multiplicative change in the hazard of construction probability; (3) values > 1 indicate higher occurrence probability relative to baseline; (4) For example, German learners (1.26) show 26.36% higher probability rates than the grand mean; (5) Two additional L1s (Turkish, Urdu) showed borderline effects with HDI bounds approaching but not reliably excluding zero, which are included in the credible count of 16 but noted as borderline; (6) Full results in Supplementary Table S2.

4.3.2. Model 2: Frequency

Model 2 examined token frequency conditional on probability, operating on 375,647 essay $\times$ verb $\times$ family observations. This model revealed the strongest L1 differentiation: 25 of 34 L1 groups (73.53%) showed credible effects.

As shown in Table 9, Arabic learners exhibited the strongest positive effect ($\beta = 0.191$, effect size = 1.21), producing approximately 21.05% more construction tokens per essay compared to the baseline. Japanese learners ($\beta = 0.104$, effect size = 1.11) and Turkish learners ($\beta = 0.097$, effect size = 1.10) also showed elevated frequencies. German learners, despite high probability, showed reduced frequency ($\beta = -0.096$, effect size = 0.91), as did Hungarian learners ($\beta = -0.113$, effect size = 0.89) and French learners ($\beta = -0.069$, effect size = 0.93). This cross-dimensional dissociation illustrates why single indices obscure meaningful variation.

Table 9

L1 groups with the largest credible effects on Model 2.

L1 background	β	94% HDI	Effect
Arabic	0.191	[0.05, 0.33]	1.21
Japanese	0.104	[0.09, 0.12]	1.11
Turkish	0.097	[0.08, 0.11]	1.10
Albanian	0.082	[0.00, 0.17]	1.09
French	-0.069	[-0.09, -0.05]	0.93
German	-0.096	[-0.11, -0.08]	0.91
Hungarian	-0.113	[-0.13, -0.10]	0.89

Note: (1) β = effect on log scale; (2) Effect = $\exp(\beta)$ = rate ratio; (3) 25/34 L1 groups showed credible effects, the strongest differentiation; (4) Full results in Supplementary Table S3.

4.3.3. Model 3: Diversity

Model 3 analyzed lexical diversity (entropy in bits) within construction families, revealing substantial cross-linguistic variation: 18 of 34 L1 groups (52.94%) showed credible effects.

As shown in Table 10, Arabic learners exhibited the highest diversity ($\beta = 0.502$), followed by German ($\beta = 0.159$) and Tswana ($\beta = 0.101$). Several L1 groups demonstrated reduced diversity, including Urdu ($\beta = -0.231$), Turkish ($\beta = -0.185$), Korean ($\beta = -0.179$), and Japanese ($\beta = -0.171$). Notably, Japanese, Korean, and Turkish learners show high frequency but low diversity, indicating an intensive yet restricted production strategy.

Table 10

L1 groups with the largest credible effects on Model 3.

L1 Background	β	94% HDI	2^β (effective-type ratio)
Arabic	0.502	[0.07, 0.92]	1.42
German	0.159	[0.11, 0.21]	1.12
Tswana	0.101	[0.05, 0.15]	1.07
Japanese	-0.171	[-0.22, -0.13]	0.89
Korean	-0.179	[-0.23, -0.13]	0.88
Turkish	-0.185	[-0.24, -0.14]	0.88
Urdu	-0.231	[-0.30, -0.16]	0.85

Note: (1) β = difference in Shannon entropy (bits) relative to the grand mean; 2^β = ratio of effective verb types (perplexity of the verb-lemma distribution) relative to baseline, so a value of

1.12 indicates a ~12% larger effective verb inventory than the grand mean; (2) 18/34 L1 groups showed credible effects; (3) Full results in Supplementary Table S4.

4.3.4. Cross-model synthesis

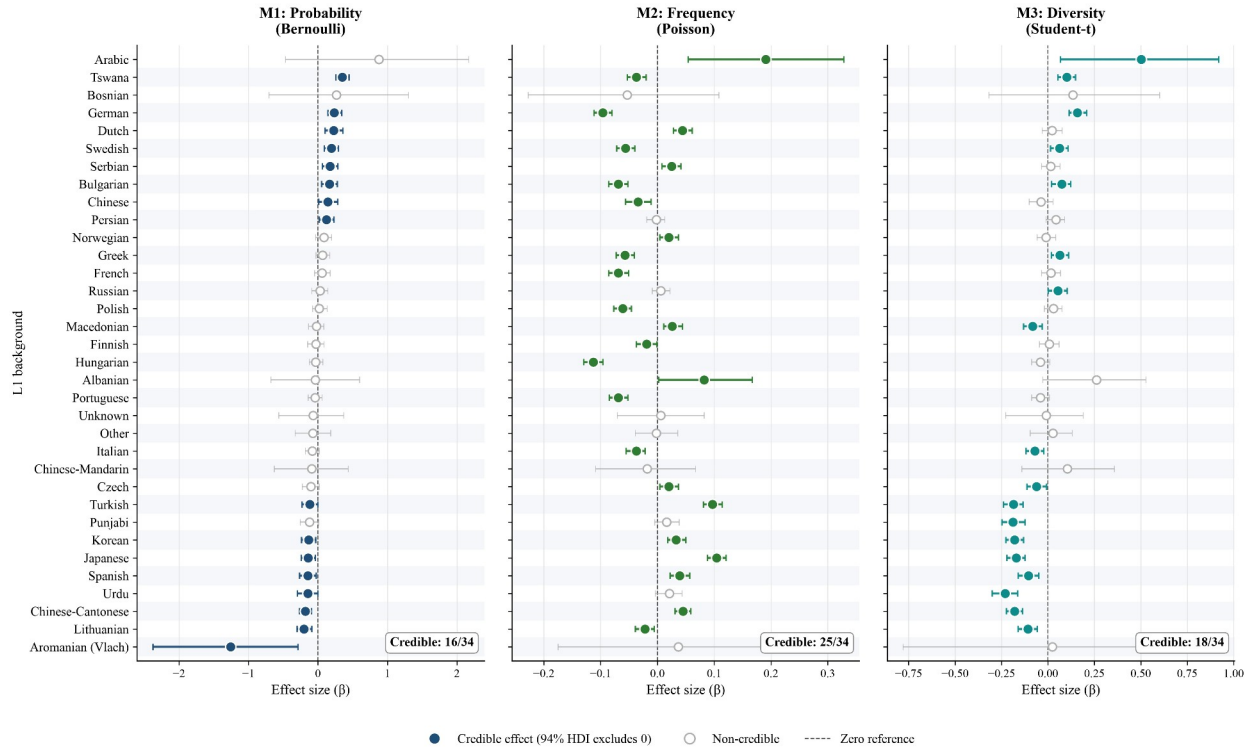


Fig. 3. L1 effects on verb-construction usage across three Bayesian models.

As shown in Figure 3, the pattern (16/34 → 25/34 → 18/34) indicates that deployment frequency is most sensitive to cross-linguistic influence, with nearly three-quarters of L1 groups (73.53%) exhibiting credible effects. In comparison, approximately 47.06% of groups show effects on probability, while about 52.94% demonstrate effects on diversity. The three models reveal that probability, frequency, and diversity represent separable dimensions that exhibit partly dissociated patterns of variation. Cross-dimensional dissociations validate the multidimensional framework. German learners exhibit high probability, low frequency, and high diversity, a selective yet lexically rich strategy. In contrast, Japanese, Korean, and Turkish learners display low probability, high frequency, and low diversity, an intensive but restricted strategy. A single composite index would obscure these distinct production profiles.

Across models, the pattern is coherent: access demonstrates moderate L1 variation (M1), deployment exhibits the strongest group-specific density preferences (M2), and lexical flexibility shows substantial variation (M3). The consistent positive effects for Arabic in M2 (21.05%) and M3 (0.50 bits, corresponding to a 1.42 effective verb inventory) are consistent with richer verb repertoires or elaborative rhetorical norms. Conversely, groups with negative effects in M3 may rely on a narrower core of high-utility verbs, resulting in efficient but less varied repertoires. It is important to emphasize that these represent production preferences, not deficits. However, the M2 frequency patterns warrant qualification. Sensitivity analyses excluding Transitive_Voice revealed that L1 frequency effects were largely driven by this high-utility family: only 6 out of 25 originally credible L1 groups retained credibility when restricted to Motion, Resultative, and Transfer

constructions (24.00% consistency). In contrast, M3 diversity effects proved robust, with 77.78% of L1 patterns persisting across the 3-family subset. This dissociation indicates that L1 influence operates differentially: affecting production frequency for high-utility constructions but primarily lexical selection for semantically constrained constructions. Full sensitivity results appear in Supplementary Materials S2.

4.3.5. Sensitivity analysis: Transitive_Voice effects on L1 differentiation

To quantify how much the Transitive_Voice family drives the L1 effects reported in Section 4.3, we re-estimated all three models restricted to Motion, Resultative, and Transfer (3-family subset). The three models diverged markedly in their robustness to this exclusion. Model 1 (probability) was the most stable: 14 of the 16 L1 groups originally credible in the full model retained credibility in the 3-family subset (87.50% consistency, all in the same direction). Model 3 (diversity) was similarly robust: 14 of the 18 originally credible L1 groups persisted with the same direction of effect (77.78% consistency). By contrast, Model 2 (frequency) was least robust: only 6 of the 25 originally credible L1 groups retained credibility in the same direction (24.00% consistency), with 19 groups losing credibility or flipping direction once Transitive_Voice was excluded.

This divergence is theoretically informative. The frequency dimension in our framework is most sensitive to L1 influence in the aggregate, but that sensitivity is disproportionately concentrated in the high-utility Transitive_Voice family, which offers learners the largest pool of voice and argument alternations to deploy. Once Transitive_Voice is removed, L1 effects on frequency are reduced to a residual core that operates on the semantically constrained Motion, Resultative, and Transfer constructions. The diversity dimension, by contrast, shows L1 effects that generalise across construction families, consistent with its interpretation as the breadth of lexical entrenchment within a construction regardless of that construction's token frequency in the corpus. Full sensitivity results for all three models appear in Supplementary Materials S2.

5. DISCUSSION

5.1. A three-stage model of L1 influence in L2 construction use

The central finding of this study is that L1 influence on argument structure construction knowledge operates differentially across three distinct dimensions. This ordering maps onto the three-component synthesis introduced in Section 1: probability reflects whether a construction crosses the activation threshold for production; frequency reflects the strength of the form-meaning mapping once it is available; and diversity reflects the breadth of lexical implementations a learner has entrenched. The finding that probability is the least L1-sensitive dimension aligns with usage-based predictions that advanced learners converge on target-language constructional inventories but retain L1-shaped preferences in how and how often those constructions are deployed (Ellis, 2006; Ellis & Sagarra, 2011). In specific, frequency showed the strongest L1 differentiation among the 34 L1 groups (25 groups, or 73.53%), followed by diversity (18 groups, or 52.94%) and probability (16 groups, or 47.06%). This ordering suggests that while basic constructional access shows moderate cross-linguistic variation at upper-intermediate to advanced proficiency, deployment choices, namely how intensively learners use constructions and with what lexical variety, remain more sensitive to cross-linguistic influence.

The pattern of results (strongest at frequency followed by diversity and probability) is consistent with a model in which L1 influence operates differentially across dimensions of constructional knowledge. While basic constructional access shows relatively limited cross-

linguistic variation at upper-intermediate to advanced proficiency, deployment preferences and lexical implementation appear more persistently shaped by L1 patterns. Whether this reflects a developmental sequence or stable individual differences requires longitudinal investigation. This aligns with usage-based predictions that deeply entrenched L1 form-meaning mappings create processing priorities that persist even when target-language constructions are formally available (Ellis, 2006; Ellis & Sagarra, 2011). The finding that frequency is more sensitive than probability to L1 influence extends this account: it suggests that L1 transfer operates not primarily at the level of grammatical availability but at the level of preferred deployment patterns, that is, when and how often to use constructions that learners demonstrably possess.

5.2. Collostructional patterns across construction families

Collostructional analysis revealed semantically coherent verb-construction associations that support the four-family taxonomy used in this study. First, Motion constructions attracted both manner verbs (e.g., *travel* $\Delta P = .87$ and *walk* $\Delta P = .85$) and the path verb *return* ($\Delta P = .80$), confirming that this family effectively captures directed motion semantics. The steep association curve indicates strong lexical entrenchment around prototypical manner verbs, with a few highly distinctive verbs followed by a gradual decline. Second, Resultative constructions favored change-of-state verbs (e.g., *keep*, *grow* and *force*) while repelling transfer verbs (e.g., *give*), aligning with the family's semantics of caused or autonomous state change. Its broader distribution reflects a wider semantic scope, covering both physical state changes and abstract result predications. Third, Transfer constructions displayed a specific profile dominated by communication verbs encoding information transfer (e.g., *show*, *offer* and *convince*), extending the prototypical transfer schema from physical objects to abstract information, as predicted by Construction Grammar (Goldberg, 1995). The gradual decay pattern suggests learners use varied communication verbs rather than relying solely on prototypical *give*. Lastly, Transitive_Voice attracted cognitive predicates (e.g., *believe* and *mean*) while repelling intransitive-preferring verbs (e.g., *fall* and *remain*), likely indicating construction boundary extensions and occasional parsing noise. The shallow cumulative coverage curve for Transitive_Voice confirms it as the broadest, most lexically distributed family.

5.3. L1-specific production strategies

The multidimensional framework reveals distinct L1-specific production strategies that single-index approaches may overlook. A key contribution of this study is the identification of cross-dimensional dissociations, showing that learners from different L1 backgrounds achieve similar overall construction use through qualitatively different profiles. German learners exemplify a broad but selective strategy: high probability ($\beta = 0.234$, effect size = 1.26) suggests productive access to all construction families; reduced frequency ($\beta = -0.096$, effect size = 0.91) indicates selective deployment; and elevated diversity ($\beta = 0.159$, $2^{\wedge}\beta = 1.12$) points to rich verb repertoires within each family. This profile reflects well-developed constructional knowledge used with restraint, indicating these learners prefer alternative syntactic strategies when constructions are not essential. In contrast, Japanese and Korean learners exhibit an intensive but restricted strategy. Reduced probability (Japanese $\beta = -0.142$, effect size = 0.87; Korean $\beta = -0.134$, effect size = 0.87) indicates narrower repertoires, while elevated frequency for Japanese ($\beta = 0.104$, effect size = 1.11) shows heavy reliance on constructions when used, and Korean displays a similar, less pronounced pattern ($\beta = 0.033$, effect size = 1.03). Reduced diversity (Japanese $\beta = -0.171$, $2^{\wedge}\beta = 0.89$; Korean $\beta = -0.179$, $2^{\wedge}\beta = 0.88$) suggests reliance on a limited set of high-utility verbs, possibly reflecting conservative strategies prioritizing accuracy or L1 patterns with

fewer verbs carrying argument structure marking. Turkish learners present a distinct restricted-usage pattern, showing borderline effects on probability ($\beta = -0.116$, effect size = 0.89) with elevated frequency ($\beta = 0.097$, effect size = 1.10) and reduced diversity ($\beta = -0.185$, $2^{\beta} = 0.88$). They produce constructions at near-baseline rates but rely intensively on a limited set of verb-construction pairings, suggesting L1 transfer of specific preferences rather than broader typological constraints. These distinct profiles highlight how aggregate measures can mask meaningful L1 variation across dimensions, which has significant implications for assessment and instruction.

6. CONCLUSION

6.1. *Major findings*

This study has three major findings. First and foremost, frequency is the most sensitive dimension to L1 background, followed by diversity and probability. This pattern reveals that basic constructional access shows moderate cross-linguistic variation at upper-intermediate to advanced proficiency, while deployment intensity and lexical implementation are more consistently influenced by L1 patterns. Secondly, frequency effects were specific to construction families, with most credible L1 frequency effects diminishing, while diversity effects remained consistent. This indicates distinct mechanisms of L1 influence on production frequency and lexical selection preferences. Thirdly, the three dimensions of probability, frequency and diversity show partly dissociated variation patterns, allowing for similar aggregate construction use through different production profiles. German learners have high probability and diversity but low frequency, while Japanese and Korean learners exhibit reduced probability and higher frequency with less diversity. Turkish learners display a borderline reduction in probability but high frequency, indicating reliance on an entrenched, narrow repertoire set of verb-construction pairings. These findings support a multidimensional approach and reveal meaningful L1 variation.

6.2. *Implications*

6.2.1. *Theoretical implications*

The staged emergence of L1 influence documented by this study carries several significant theoretical implications for our understanding of language acquisition and production. First, $M1 = 16/34$ indicates that nearly half of L1 groups show credible effects on construction probability, though this represents the weakest dimension of L1 influence. This finding suggests that while basic constructional access shows moderate cross-linguistic variation at upper-intermediate to advanced proficiency, most learners have acquired productive access to argument structure patterns. Second, $M2 = 25/34$ reveals that the extent to which learners use each construction varies significantly by L1, indicating L1-specific preferences in production frequency. Third, $M3 = 18/34$ shows that lexical diversity within constructions also varies according to L1, suggesting differences in the depth of learners' repertoires. To sum up, L1 background moderately affects whether learners *use* a construction (probability: M1), strongly influences how *often* they use it (frequency: M2), and substantially affects how *diversely* they employ it (diversity: M3). This supports a usage-based interpretation: all learners eventually encounter and produce the same construction types, but their frequency and lexical diversity patterns differ based on L1 transfer effects. This finding aligns with research showing that L1 effects are more subtle than simple presence/absence: they manifest in how constructions are used, not whether they are used (Ellis, 2006; Ellis & Sagarra, 2011).

6.2.2. Methodological implications

Methodologically, this study illustrates the significant advantages of utilizing Bayesian hierarchical modeling in learner corpus research. By shifting focus from aggregated frequency counts to principled statistical inference, this approach effectively addresses the nested structure of corpus data, allowing for a more sophisticated analysis. Distinctly modeling probability, frequency, and diversity as separate outcomes prevents the conflation of access with deployment preferences, an issue prevalent in prior studies that relied on aggregate token counts. This collapse of dimensions can obscure nuanced staged L1 effects. Thus, the multidimensional framework established in this research lays a solid foundation for more detailed investigations into L2 constructional development, facilitating a deeper understanding of learners' language acquisition processes.

Moreover, collostructional analysis uncovered semantically coherent verb-construction associations that support the validity of the four-family taxonomy: Motion, Resultative, Transfer, and Transitive_Voice. The strong attraction of manner-of-motion verbs to Motion constructions, change-of-state verbs to Resultative constructions, and communication verbs to Transfer constructions confirms that the automated annotation pipeline effectively identified instances of constructions with the appropriate semantic properties. This alignment between verb types and their corresponding constructions not only reinforces the theoretical framework but also demonstrates the utility of the analytical methods employed in discerning meaningful patterns within learner corpora.

6.2.3. Pedagogical implications

The dimensional framework provides significant diagnostic potential for second language assessment and targeted instruction by emphasizing the multifaceted nature of construction knowledge. Instead of viewing this knowledge as a singular entity, assessment tools can be designed to evaluate distinct dimensions such as probability, frequency, and diversity of constructions. This nuanced approach recognizes that learners with similar overall scores on assessments may possess different underlying profiles across these dimensions, which would necessitate tailored instructional interventions to meet their specific needs.

For instance, German learners typically demonstrate high diversity in their use of language constructions but may benefit more from strategies aimed at enhancing their production frequency. In contrast, learners from Japanese, Korean, and Turkish backgrounds might find it advantageous to focus on broadening their verb repertoires within various constructions. This shift involves moving beyond commonly used, high-utility verbs to embrace a wider array of lexical choices, thereby enriching their expressive capabilities.

By making L1-specific adjustments to instruction based on the patterns observed in this study, educators can create more effective learning experiences. Ultimately, validating these targeted instructional strategies through intervention studies could lead to a more efficient allocation of pedagogical resources compared to traditional undifferentiated teaching methods, ensuring that all learners receive the support they need to develop their L2 skills effectively.

6.3. Limitations and future directions

Several limitations warrant acknowledgment. First, the cross-sectional design prevents conclusions about developmental trajectories; the dimensional patterns observed may not persist at other proficiency levels. Second, typological categories were not directly modeled as predictors. Third, the absence of a native speaker baseline limits interpretation of whether observed L1

patterns represent approximations toward, or divergences from, target norms. Fourth, the aggregation of active transitive and passive realizations into a single Transitive_Voice family preserves statistical power at the cost of masking voice-specific transfer patterns.

Future studies can work in the following directions. First, longitudinal studies would establish whether dimensional dissociations reflect stable L1 influence or proficiency-dependent states. Second, future work could include formal typological features (e.g., satellite-framed vs. verb-framed classification) as predictors, enabling direct evaluation of whether typological categories explain variance beyond individual L1 effects. Third, future work comparing learner profiles against native benchmarks would clarify the acquisitional significance of these patterns. Last, future work should model TRAN_S and PASSIVE as separate families, ideally with native-speaker baselines for voice preference across typologically diverse L1s, so that passive-specific L1 effects can be distinguished from active-transitive effects.

References

- Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesebeck, C. J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C. C., Martin, O. A., Osthege, M., Vieira, R., Wiecki, T., & Zinkov, R. (2023). PyMC: A modern, and comprehensive probabilistic programming framework in Python. *PeerJ Computer Science*, 9, e1516.
- Aljadani, M. (2018). Second language acquisition of the English dative alternation by native speakers of Arabic. *Journal of Language and Cultural Education*, 6(1), 65–78.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Boyd, J. K., & Goldberg, A. E. (2009). Input effects within a constructionist framework. *The Modern Language Journal*, 93(3), 418–429. <https://doi.org/10.1111/j.1540-4781.2009.00899.x>
- Bozdağ, F. (2024). Probabilistic analysis of English dative constructions in academic writings of English EFL learners. *Theory and Practice of Second Language Acquisition*, 10(1), 1–24. <https://doi.org/10.31261/tapsla.13902>
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.
- Cadierno, T. (2008). Learning to talk about motion in a foreign language. In P. Robinson & N. C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 239–275). Routledge.
- Cadierno, T., & Ruiz, L. (2006). Motion events in Spanish L2 acquisition. *Annual Review of Cognitive Linguistics*, 4, 183–216. <https://doi.org/10.1075/arcl.4.08cad>
- Callies, M., & Szczesniak, K. (2008). Argument realisation, information status and syntactic weight – A learner-corpus study of the dative alternation. In M. Walter & P. Grommes (Eds.), *Fortgeschrittene Lernervarietäten* (pp. 165–187). Niemeyer.
- Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3), 369–382. <https://doi.org/10.1177/0267658312443651>
- Diessel, H. (2019). *The grammar network: How linguistic structure is shaped by language use*. Cambridge University Press. <https://doi.org/10.1017/9781108671040>
- Divjak, D. (2019). *Frequency in language: Memory, attention and learning*. Cambridge University Press. <https://doi.org/10.1017/9781316084410>

- 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. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, 27(2), 164–194. <https://doi.org/10.1093/applin/aml015>
- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17–44. <https://doi.org/10.1017/S0267190512000025>
- Ellis, N. C., & Ferreira-Junior, F. (2009). Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal*, 93(3), 370–385. <https://doi.org/10.1111/j.1540-4781.2009.00896.x>
- Ellis, N. C., & Sagarra, N. (2011). Learned attention in adult language acquisition: A replication and generalization study and meta-analysis. *Studies in Second Language Acquisition*, 33(4), 589–624. <https://doi.org/10.1017/S0272263111000325>
- Eskildsen, S. W. (2012). L2 negation constructions at work. *Language Learning*, 62(2), 335–372. <https://doi.org/10.1111/j.1467-9922.2012.00698.x>
- Filipović, L., & Hawkins, J. A. (2013). Multiple factors in second language acquisition: The CASP model. *Linguistics*, 51(1), 145–176. <https://doi.org/10.1515/ling-2013-0005>
- Gedik, T. A., & Uslu, F. (2023). L1-L2 transfer in ditransitive construction: A usage-based replication study with Turkish speakers of English. In G. P. Georgiou, A. Giannakou, & C. Chrysochoou (Eds.), *Advances in Second/Foreign Language Acquisition* (pp. 123–146). Springer International Publishing. https://doi.org/10.1007/978-3-031-38522-3_6
- Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.
- Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press.
- Goldberg, A. E., Casenhiser, D., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, 15(3), 289–316. <https://doi.org/10.1515/cogl.2004.011>
- Goschler, J., & Stefanowitsch, A. (2024). Constructional transfer in L2 learning: The role of entrenchment. *Zeitschrift für Anglistik und Amerikanistik*, 72(3), 283–317. <https://doi.org/10.1515/zaa-2024-2028>
- Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2020). *International Corpus of Learner English* (Version 3). Presses Universitaires de Louvain.
- Gries, S. T. (2015). The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora*, 10(1), 95–125. <https://doi.org/10.3366/cor.2015.0068>
- Gries, S. T. (2018). On over- and underuse in learner corpus research and multifactoriality in corpus linguistics more generally. *Journal of Second Language Studies*, 1(2), 276–308. <https://doi.org/10.1075/jsls.00005.gri>
- Gries, S. T. (2021). *Statistics for Linguistics with R: A Practical Introduction* (3rd ed.). De Gruyter Mouton.
- Gries, S. T., & Wulff, S. (2005). Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics*, 3(1), 182–200. <https://doi.org/10.1075/arcl.3.10gri>

- Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15, 1593–1623.
- Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength natural language processing in Python. Zenodo. <https://doi.org/10.5281/zenodo.1212303>
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461–473. <https://doi.org/10.1093/applin/amp048>
- Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. John Benjamins. <https://doi.org/10.1075/llt.32>
- Hwang, H., & Kim, H. (2022). Automatic analysis of constructional diversity as a predictor of EFL students' writing proficiency. *Applied Linguistics*, 44(1), 127–147. <https://doi.org/10.1093/applin/amac046>
- Ibarretxe-Antuñano, I. (2009). Path salience in motion events. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & S. Özçalışkan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 403–414). Psychology Press.
- Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge. <https://doi.org/10.4324/9780203935927>
- Kim, H., & Shin, G.-H. (2025). The interplay of verbs and argument structure constructions in second language processing: Roles of verb's lexical properties and verb–construction association. *Cognitive Linguistics*, 36(1), 59–88. <https://doi.org/10.1515/cog-2024-0042>
- Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, 1(2), 270–280. <https://doi.org/10.1177/2515245918771304>
- Kyle, K. (2016). *Measuring Syntactic Development in L2 Writing: Fine Grained Indices of Syntactic Complexity and Usage-based Indices of Syntactic Sophistication*. Unpublished doctoral dissertation, Georgia State University.
- Langacker, R. W. (1987). *Foundations of cognitive grammar: Vol. 1. Theoretical prerequisites*. Stanford University Press.
- Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press.
- Liu, Y., & Lu, X. (2024). Development of verb argument constructions in L2 English learners: A close replication of research question 3 in Römer and Berger (2019). *Studies in Second Language Acquisition*, 46(5), 1373–1391. <https://doi.org/10.1017/S027226312400024X>
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). CRC Press.
- Norouzian, R., de Miranda, M., & Plonsky, L. (2018). The Bayesian revolution in second language research: An applied approach. *Language Learning*, 68(4), 1032–1075. <https://doi.org/10.1111/lang.12310>
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555–578. <https://doi.org/10.1093/applin/amp044>
- Römer, U., & Berger, C. M. (2019). Observing the emergence of constructional knowledge: Verb patterns in German and Spanish learners of English at different proficiency levels. *Studies in Second Language Acquisition*, 41(5), 1089–1110. <https://doi.org/10.1017/S0272263119000202>

- Römer, U., O'Donnell, M. B., & Ellis, N. C. (2014). Second language learner knowledge of verb-argument constructions: Effects of language transfer and typology. *The Modern Language Journal*, 98(4), 952–975. <https://doi.org/10.1111/modl.12149>
- Schmid, H.-J. (2017). A framework for understanding linguistic entrenchment and its psychological foundations. In H.-J. Schmid (Ed.), *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge* (pp. 9–35). American Psychological Association.
- Schmid, H.-J. (2020). *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment*. Oxford University Press.
- Slobin, D. I. (1996). From 'thought and language' to 'thinking for speaking.' In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking Linguistic Relativity* (pp. 70–96). Cambridge University Press.
- Slobin, D. I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömquist & L. Verhoeven (Eds.), *Relating events in narrative: Typological and contextual perspectives* (pp. 219–257). Lawrence Erlbaum Associates.
- Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243. <https://doi.org/10.1075/ijcl.8.2.03ste>
- Sung, H., & Kyle, K. (2024). Leveraging pre-trained language models for linguistic analysis: A case of argument structure constructions. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 7302–7314). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.415>
- Talmy, L. (2000). *Toward a cognitive semantics* (Vols. 1–2). MIT Press.
- Tomasello, M. (2003). *Constructing A Language: A Usage-based Theory of Language Acquisition*. Harvard University Press.
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*, 16(2), 667–718. <https://doi.org/10.1214/20-BA1221>
- Verspoor, M., Schmid, M. S., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, 21(3), 239–263. <https://doi.org/10.1016/j.jslw.2012.03.007>
- Year, J., & Gordon, P. (2009). Korean speakers' acquisition of the English ditransitive construction: The role of verb prototype, input distribution, and frequency. *The Modern Language Journal*, 93(3), 399–417. <https://doi.org/10.1111/j.1540-4781.2009.00898.x>