LingBaW. Linguistics Beyond and Within 11 (2025), 69–88

Teaching Language-Specific Preferences in the Native Language: An Interactive Web App for Prospective Teachers

Renate Delucchi Danhier[1] and Paula González Ávalos[2]
[1]TU Dortmund University, Germany; [2]appliedAI Institute for Europe, Germany

Abstract

Natural languages offer diverse means for conveying the same idea, with each language typically exhibiting a predominant way of selecting and packaging information, referred to as a language-specific preference. Cross-linguistic research on motion events has shown that native speakers tend to favor such patterns and intuitively recognize non-preferred lexicalizations, particularly in the domain of motion event description. Due to subtle differences in grammar and vocabulary, these preferences are language-specific.

Acquiring the language-specific preferences of a second or foreign language poses a challenge even for advanced learners, especially when these preferences diverge from those internalized in the native language. Language teachers, however, are often unaware of both the existence of such preferences and the difficulties they pose for learners. This study therefore aims (1) to enable prospective teachers to discover language-specific preferences through a bottom-up, data-driven approach, and (2) to raise awareness of the challenges involved in acquiring them.

Accordingly, the online application *LexiGraph* was developed to enable prospective German teachers to identify patterns in linguistic data. Using an interactive dataset, students explored the expression of motion events across languages and speaker groups. The app supports visual exploration and provides a threshold learning experience that fosters an engaging and intellectually challenging learning environment by encoding complex semantic and syntactic information.

Student feedback highlights the novelty of the technological approach, although some participants perceived the material as cognitively demanding. Developing an understanding of language-specific lexicalization patterns can support prospective teachers in working with language learners and contribute to more inclusive multilingual education.

Keywords: teacher training; motion event construal; web application; language awareness competence; digital humanities

LingBaW. Linguistics Beyond and Within, e-ISSN: 2450-5188
DOI: 10.31743/lingbaw.19443
© The Author(s) 2025. This is an open access article licensed
under a Creative Commons Attribution 4.0 International License.

1. Introduction

When talking about motion in space in any natural language, a specific situation can be expressed in various ways. However, the majority of speakers of a given language tend to encode the same information using recurring linguistic means, consistently selecting the same form–meaning constructions. This preference seems to arise because a particular way of packaging information aligns well with the grammar and lexicon of that language. Since grammatical and lexical systems differ across languages, so do optimal packaging strategies, giving rise to language-specific preferences in motion event construal.

Acquiring a language does not only involve learning its grammar and vocabulary, but also developing language-specific modes of thinking. The concept of *thinking-for-speaking* (Slobin 1991, 1996a) refers to the particular kind of thinking that takes place just before speaking, when speakers plan how to express information in a specific language. As Slobin (1996a: 89) notes, "each native language has trained its speakers to pay different kinds of attention to events and experiences when talking about them", depending on which features are grammaticalized in that language.

Learning the language-specific preferences of a target language that differ from those of one's native language constitutes a linguistic challenge even for advanced second-language users. Such L2 speakers often produce grammatically correct utterances but fail to conform to the preferred patterns of the target language. Instead, they frequently retain patterns from their first language, resulting in negative transfer (Alghamdi, Daller & Milton 2019; Antonijević & Berthaud 2009). These non-preferred constructions, while formally still grammatically correct, often sound peculiar to monolingual native speakers. Bilingual speakers may also produce utterances perceived as unusual by monolinguals, particularly in their non-dominant language. Because divergent preferences are not necessarily ungrammatical but merely less frequent, advanced learners rarely receive corrective feedback from teachers or other speakers.

In school contexts, bilingual pupils and advanced second-language learners would therefore benefit from teachers who understand language-specific preferences and can make them explicit through instruction. Unfortunately, most trained teachers remain unaware of the complexity of cross-linguistic differences in the domain of motion, leaving learners to navigate these subtleties largely on their own.

Although difficulties with motion verbs in second language acquisition have been documented for decades (e.g., Schlyter 1984; Schlyter & Viberg 1985), these insights have largely remained confined to research contexts and have rarely been integrated into teacher education or classroom practice. Schlyter's (1984) work on Swedish learners of French and Schlyter and Viberg's (1985) comparative study of French and Swedish show that learners struggle to acquire the lexical and semantic distinctions required to encode motion appropriately in an L2. Their findings point to typological and cross-linguistic influences that affect how motion is verbalized.

Our objective is therefore not to claim novelty in identifying this linguistic challenge, but to address its pedagogical invisibility. We aim to make prospective teachers aware of cross-linguistic differences in motion event construal, arguing that increased awareness of this

persistent learning barrier can enhance their ability to empathize with and support their future pupils.

Our pedagogical framework draws on the concept of *threshold learning* (Meyer & Land 2005; Land et al. 2010), which refers to transformative moments in which learners acquire new understandings that fundamentally change their way of thinking about a subject. Threshold concepts are often difficult to grasp and emotionally demanding, because they challenge prior assumptions and require a shift in perspective. Once internalized, however, they bring about irreversible cognitive and ontological change: learners not only acquire new knowledge, but also come to see the discipline differently. In language education, such concepts often involve recognizing hidden, language-specific constraints that native speakers apply unconsciously.

The *LexiGraph* application was designed to provoke precisely this kind of threshold learning experience. By confronting students with visually encoded data derived from authentic motion event descriptions, the interface initially appears complex and even disorienting. This deliberate challenge invites learners to explore the data, formulate hypotheses, and gradually uncover underlying linguistic patterns. The moment when the system's visual logic becomes transparent parallels the conceptual breakthrough central to threshold learning: implicit linguistic knowledge is transformed into explicit awareness of how languages differ in mapping meaning onto form. Moreover, the process of discovering hidden regularities in the corpus mirrors the inductive nature of first language acquisition. By engaging with this experience, prospective teachers (who completed their own language acquisition process long ago) are reminded of how demanding it is for learners to infer such subtleties from input alone, without explicit instruction.

2. Differences in the construal of motion events across languages

Research on motion events investigates how languages encode movement through grammatical and lexical means. In a motion event, an entity (the FIGURE) moves relative to a reference point (the GROUND), typically along a TRAJECTORY that begins at a SOURCE and ends at a GOAL. Cognitive linguistics models this structure using the Source–Path–Goal schema (Lakoff 1987), emphasizing the link between conceptualization and linguistic expression. Figure 1 illustrates how speakers select and organize information differently when describing the same situation.

A man is walking down the street.
Someone is taking a walk in the city.
A person is walking past a bus station.
A man is walking to the supermarket.
A man has just left the house.
Someone is going shopping.
Someone is beginning their day.
A man is hurrying down the street in a city.
I see a city landscape.
A man runs along an empty freeway.
He is probably heading towards the shop.

Figure 1: *Given the same situation, speakers have various options for selecting and organizing information for expression*

Consider the English sentence:

(1) A man walks from his home across the street to the supermarket.

FIGURE			GROUND		GROUND		GROUND
A man	walks	from	his home	across	the street	to	the supermarket
	MANNER	SOURCE		TRAJECTORY		GOAL	

Here, *walk* encodes the MANNER of motion, while the prepositional phrases express the SOURCE, TRAJECTORY, and GOAL respectively. Languages differ in how they distribute semantic components between the verb and elements outside the verb, commonly referred to as satellites.

All examples of verbalization presented in Figure 1 capture the situation equally well; however, some sound more natural or less forced than others to native speakers of English. The sentences are ordered from top to bottom according to approximate frequency of use, ranging from most to least frequent, based on language production experiments (Stutterheim et al. 2012). To understand why some constructions sound less natural than others despite being grammatically correct, it is necessary to examine how lexis and grammar interact in the formation of complex constructions.

2.1. Lexicalization patterns

Languages vary systematically in how they integrate manner and path information into motion descriptions. Some languages favor verbs that encode manner (how something moves), while others rely primarily on verbs expressing path (direction or trajectory). German, like English, possesses a rich vocabulary of manner-of-motion verbs, such as *laufen, eilen, rennen, kriechen, schlendern, gleiten, stolzieren, stapfen, taumeln*, and *waten*, which parallel English *walk, dash, run, crawl, stroll, slide, strut, trudge, stagger*, and *wade*.

When speakers wish to convey several semantic components within a single sentence (as shown in Figure 1), only some of these can be encoded in the verb itself (often only one), while other information must be encoded elsewhere (outside the verb), for instance through prepositional phrases or adverbs. The systematic pairing of form and meaning used to express

complex information within a sentence, while respecting the lexical and grammatical constraints of a language, is called a *lexicalization pattern*.

Following Talmy's (2000) typology, German and English are classified as satellite-framed languages, meaning that the verb typically encodes manner, while path is expressed in an accompanying satellite such as a prepositional phrase or verb particle (*hinauslaufen, in das Haus gehen* / *run out, go into*). In contrast, verb-framed languages (e.g., Spanish and French) encode path in the verb (*entrar*, *salir*) and use additional elements for manner, if at all. English occupies an intermediate position: while it largely behaves as a satellite-framed language, it also contains path verbs of Romance origin (*enter, exit, ascend, descend*), which do not readily combine with satellites expressing manner (*?She entered the cave running*). German, lacking this Romance influence, relies more consistently on manner verbs combined with particles or prepositions (e.g., *Der Wolf läuft aus dem Wald hinaus*). Subsequent work has refined Talmy's binary typology to account for additional patterns observed cross-linguistically. Slobin (2004) and Chen and Guo (2009) proposed a third type, equipollently-framed languages, such as Chinese, in which both the verb and its accompanying element (often a directional complement) contribute equally to encoding manner and path (E-framing). Moreover, double-marking constructions (Bohnemeyer et al. 2007; Croft et al. 2010) describe cases in which both the verb and a satellite simultaneously encode path or direction (also referred to as double-framing). These refinements acknowledge that languages often combine multiple strategies rather than adhering strictly to a single framing type.

Cross-linguistic research has shown that such lexicalization patterns reflect language-specific preferences rather than absolute constraints (Slobin 2004; Croft et al. 2010). Both English and German employ multiple construction types; however, English tends to use satellite structures more freely and often focuses on manner, whereas German shows greater flexibility in encoding path through prefixes or compounds (e.g., *hinaus-, hinein-*). As a result, speakers of each language develop distinct intuitions about which combinations sound natural, even when several options are grammatically correct.

Interlanguage research provides further evidence for the persistence of L1-based lexicalization patterns in second-language learners. Hill (1992), analyzing the mental lexicon of Kenyan learners of English motion verbs, found that learners tended to group verbs according to path rather than manner, reflecting L1-driven semantic organization. Using cluster analysis, Hill demonstrated that interlanguage lexicons reveal systematic differences in conceptual grouping, confirming that typological contrasts between languages extend to the mental organization of motion semantics.

2.2. Grammatical aspectual perspectivation

Beyond lexicalization, grammatical aspect influences how speakers conceptualize motion. English distinguishes between ongoing (*was running*) and completed (*ran*) motion, allowing speakers to highlight either the process or the endpoint. German, which lacks a grammaticalized aspect system, primarily conveys such distinctions through lexical or contextual means.

Comparative studies of aspectual and non-aspectual languages (von Stutterheim & Nüse 2003; Carroll et al. 2004) show that speakers of languages with progressive aspect (like English) tend to focus on the process of movement, while speakers of languages without progressive aspect (like German) more often include goal information, even when goal attainment is uncertain (Figure 2). Although it is perfectly grammatical in English to specify a potential goal (*A man is going to the supermarket*), only a minority of speakers actually choose to do so. These differences in perspective affect both linguistic and non-linguistic behavior, including event recall (memory tests) and visual attention as measured through eye-tracking (Bylund 2008; Stutterheim et al. 2012; Athanasopoulos & Bylund 2013). As a result, these effects are assumed to have a cognitive dimension rooted in grammatical distinctions between languages (Schmiedtová et al. 2011).

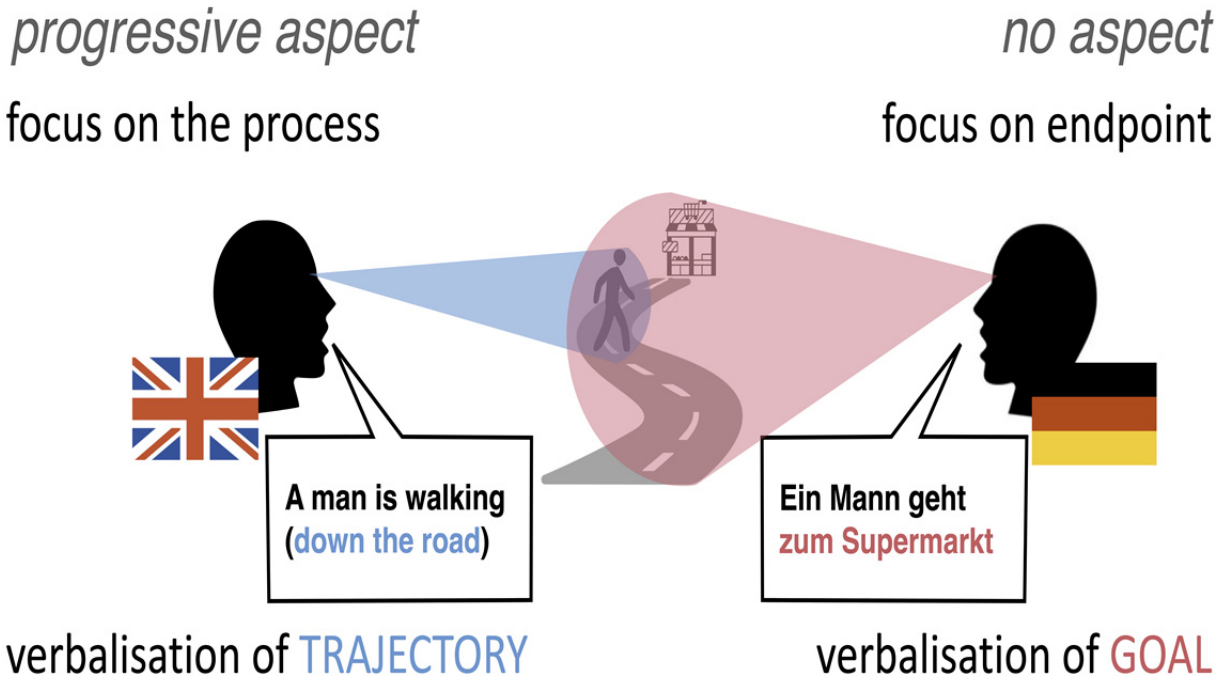

Figure 2: *Depending on whether a language has a progressive aspect, native speakers adopt different perspectives on motion events, leading to the selection of different information for verbalization*

2.3. Implications for cross-linguistic comparison

Typological classifications should therefore be understood as tendencies rather than absolute distinctions (Berman & Slobin 1994; Talmy 2000). Although both English and German display intra-linguistic variation in motion event encoding, they differ systematically in the balance between manner and path information: English strongly favors satellite-framed constructions and manner verbs, whereas German, despite many similarities, tends to foreground path and goal components slightly more consistently (Slobin 1996b; Montero-Melis et al. 2017). Recognizing these subtle yet systematic differences provides an essential foundation for training language teachers to identify implicit patterns in how their native language structures meaning.

In cross-linguistic research, it is crucial to focus on elements that are genuinely comparable. Typological comparisons such as those facilitated by our app do not examine entire language systems; instead, they analyze constructions used by real speakers in specific contexts to convey equivalent meanings (Croft 2003:13). In the domain of motion, this entails comparing how speakers of different languages verbalize the same situations by examining the morphosyntactic forms employed and the meanings selected. These pairings of forms and meanings are termed constructions (Croft 2001), and range from individual morphemes to complex clauses. The fundamental units of comparison are thus the constructions employed in various languages to encode specific situation types. It is important to briefly acknowledge that this perspective differs from the initial typological approach that aimed to classify entire language systems rather than specific construction types. Berman and Slobin (1994: 118) noted limitations of this approach, cautioning that "as a general caveat, it should be remembered that typological characterizations often reflect tendencies rather than absolute differences between languages." Talmy (2000:65) similarly notes that languages may employ different conflation types for different kinds of motion events or offer multiple options for the same type: "... a language can characteristically employ one conflation type for one type of motion event and characteristically employ a different conflation type for another type of motion event".

In our perspective, the widespread intra-linguistic variation in motion event encoding should not be treated as mere exceptions: Variations within languages emerge not only in the selection of meaning for expression, but also in the forms used to encode specific meanings. Satellite-framed languages often exhibit a path bias, emphasizing path information, sometimes in particular goal information, more frequently, whereas verb-framed languages more commonly encode such information in the main verb (Slobin 1996b; Johanson & Papafragou 2010). In certain situation types, such as downward motion or unreached goals, path or goal prominence is especially pronounced in satellite-framed languages (Slobin 1996b:199).

Speakers of non-aspectual languages commonly emphasize the goal in verbalization, while speakers of aspectual languages rarely do so (von Stutterheim et al. 2012). English additionally shows a tendency to encode path information in satellites. These linguistic differences in how spatial information is packaged can be attributed to variations in lexicon and grammar. Decades of comparative linguistic research on motion event verbalization have identified both language-specific obligatory restrictions and complex non-obligatory preferences.

Lexicalization patterns and grammatical aspectual perspectivation thus reveal systematic ways in which spatial concepts are linguistically encoded, offering insights into cognitive, cultural, and grammatical influences on language use. These patterns inform how information is structured and prioritized, providing valuable insights into how speakers conceptualize and communicate ideas in their language according to preferences they have been learning for a lifetime from the input. While expecting a language to exhibit only a single lexicalization pattern or to obligatorily express specific spatial information would be an oversimplification, such generalizations can nevertheless serve pedagogical purposes.

The question therefore arises whether language teachers, equipped with native intuitions in their first language and explicit knowledge of a second language, are able to rediscover

these intricacies, or (failing that) to appreciate the magnitude of the challenge faced by language learners.

3. State of the art: Teaching lexicalization patterns

Decades of psycholinguistic research have examined how languages construe motion events, identifying systematic patterns and tracing their acquisition processes (Berman & Slobin 1994). Numerous studies have examined the varying degrees of success with which second-language speakers acquire these patterns (Cadierno 2004; Navarro & Nicoladis 2005; Cadierno & Ruiz 2006). Factors such as event segmentation and endpoint orientation have also been shown to influence adaptation in both second language acquisition (Athanasopoulos et al. 2015) and foreign language learning (Schmiedtová & Flecken 2008).

Despite this extensive body of research, comparatively little attention has been devoted to pedagogical strategies for teaching lexicalization patterns. Early theoretical proposals advocate a focus-on-form approach to grammar instruction (Cadierno 2008). However, several challenges complicate the teaching of motion event construal: framing patterns are rarely emphasized (sometimes entirely neglected) in grammar classes (Cadierno & Robinson 2009); learners receive limited exposure to target constructions in classroom settings and therefore lack positive evidence (Treffers-Daller & Tidball 2015); and, most importantly for the present project, teachers themselves often remain unaware of the cross-linguistic complexity of the motion domain (Attwood 2014).

More recent pedagogical research has begun to address these challenges. In an intervention study, Spanish speakers learning English successfully "unlearned" the boundary-crossing constraint through targeted training. After instruction, they showed a preference for satellite-framed constructions in English and generalized this preference to non-boundary-crossing contexts (Laws, Attwood & Treffers-Daller 2021). A related constructivist initiative closely aligned with the present project involved a digital escape room that enabled Spanish learners of English to discover the contrast between satellite-framed and verb-framed constructions through guided exploration (Gutiérrez & Costa 2021). This learner-centered approach resonates with our aim of allowing participants to explore framing patterns autonomously through visual representations of grammatical structure.

Visual-based grammar instruction has a long tradition in language education, ranging from concrete aids such as images that enhance engagement to more abstract visualizations of grammatical concepts, for example graphical representations of English tense systems (Beare 2015). Techniques such as visual input enhancement (Park et al. 2012), i.e., the deliberate use of color, boldface, or capitalization to highlight grammatical features in textbook materials, have also been applied to the teaching of motion event construal (Attwood 2014). Taken together, these studies demonstrate that visual design can effectively guide learners' attention toward language-specific lexicalization patterns, providing a strong foundation for the pedagogical approach developed in the present project.

4. Methods

The proposed approach to linguistic training establishes a threshold learning environment that emulates the pattern-recognition mechanisms employed in first language acquisition, while employing a novel visual system for encoding linguistic information. The design of the application deliberately avoids simplifying the complex interaction between semantics and syntax that characterizes the verbalization of motion events.

4.1. Technical specifications and functions of the app

The proposed solution is the web-based application *LexiGraph*, designed to run in a standard web browser. *LexiGraph* provides a visual representation of lexicalization patterns through complex graphics that combine spatial concepts with syntactic means. These visual representations are inspired by idiographic writing systems, such as logographic scripts including Chinese *Hànzì* or ancient Egyptian hieroglyphs, in which individual characters encode semantic components or morphemic information.

The app processes a dataset consisting of a table of manually coded linguistic data. It was developed in R using the Shiny framework (R Core Team 2022), which enables the creation of interactive web applications directly from R code. Shiny integrates the statistical and graphical capabilities of R with a dynamic user interface, allowing users (in this case, students) to manipulate variables and visualize results in real time without requiring programming knowledge. This makes the framework particularly suitable for teaching contexts in which learners explore linguistic data through interactive visual representations rather than static charts. The application was developed using the packages shinyWidgets, tidyverse, and ggimage and comprises 248 lines of code. It is accessible online (Delucchi Danhier & González Ávalos, *LexiGraph*).

The user interface (Figure 3) consists of interactive panels that allow students to filter and compare verbalizations according to selected linguistic features, such as figure, ground, trajectory, or goal. Users can toggle categories, switch between speaker groups (e.g., German monolingual vs. English L2 speakers), or stimulus sets, and view the visualizations of the verbalizations of two groups side by side in real time. This interactivity transforms the dataset into an exploratory learning environment: rather than passively reading about cross-linguistic preferences, students can observe how different languages encode motion events and how these patterns vary across speakers and stimuli. The visual feedback supports hypothesis testing and pattern recognition, two processes central to the discovery-based learning approach underpinning the project. The design of the user interface aims to create a threshold learning experience: The initial encounter with the visual system may feel demanding or even frustrating, mirroring aspects of language learning itself. This challenge is deliberate, as it encourages learners to engage deeply with the material and to experience first-hand the type of conceptual shift they are expected to understand and later facilitate in their own teaching.

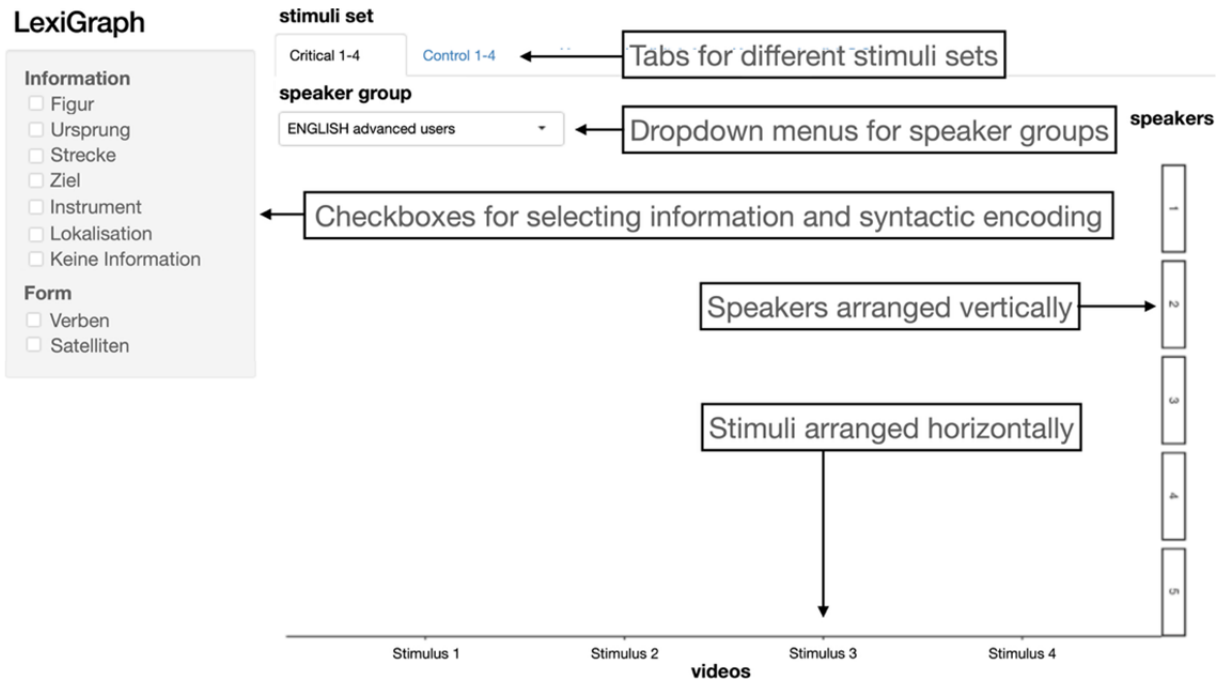

Figure 3: *User interface of LexiGraph; selection panels at the top and left allow users to filter subsets of the dataset and focus on specific aspects of interest*

Students first familiarized themselves with the visual language of the application by working in pairs and subsequently explored the dataset collaboratively in small groups, searching for patterns in the data. By alternating between empirical data exploration and theoretical discussion, including engagement with specialized literature, students developed an understanding of language-specific lexicalization patterns and preferences in German through a bottom-up approach.

4.2. Data collection and preparation

The data used in the application were collected by the students themselves. The experimental design closely follows that employed by Schmiedtová et al. (2011) and von Stutterheim et al. (2012), albeit in an abridged form, with fewer scenes and no memory task. This design made it possible to collect oral data on motion event descriptions. The stimulus set consisted of 16 short video clips depicting everyday situations. Four critical scenes showed goal-oriented locomotion events in which the video ended before the potential goal was reached. Four control scenes depicted goal-oriented locomotion events in which the goal was clearly reached. The remaining eight scenes served as distractors and did not involve locomotion. The design was based on the assumption that speakers would mention the goal more frequently in control scenes than in critical scenes. The video clips used have been employed in motion event research for many years (Figure 4). Stimuli were presented in a pseudo-randomized order to prevent critical and control scenes from appearing consecutively. Participants received the following instruction: *Say what is happening in each video. Do not describe, just focus on the action.* The corresponding German instruction was: *Sagen Sie, was in jedem Video passiert. Beschreiben Sie bitte nicht, sondern konzentrieren Sie sich auf die Handlung.*

Figure 4: *Examples of video stimuli showing start and end frames; in critical scenes, the protagonist's arrival at the goal is uncertain, while in control scenes the goal is clearly reached*

Participants were classified into a monolingual German speaker group and a very advanced learner-of-English group on the basis of a short language biography questionnaire. Because the German school curriculum includes English classes starting in fifth grade, it is nearly impossible to find completely monolingual university students in Germany. In the context of this study, the term *monolingual* refers to individuals who grew up speaking only German at home and reported minimal proficiency in other languages. Students of English studies completed the task in English, whereas monolingual German speakers completed it in German. Students who did not fit either category were asked to recruit a person from their social network who had either grown up monolingually or used English professionally. All audio recordings were transcribed, segmented into utterances, and subsequently coded by a comparative linguist with expertise in lexicalization patterns.

4.3. Coding and visualization of the linguistic data

The coding scheme follows the spatial categories proposed by Delucchi Danhier (2017), although the visual language of the application uses updated colors and symbols, as shown in Figure 5. The coded meanings include spatial concepts such as deictic orientation, location, and instrument, which are semantic components not addressed in Talmy's original framework.

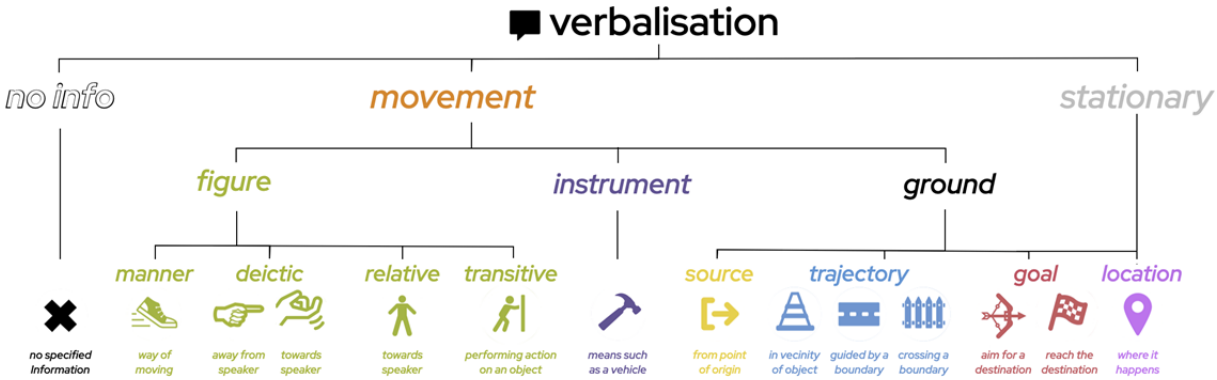

Figure 5: *Colors and symbols used to represent the coded spatial categories*

The visually encoded data capture the number of utterances used to verbalize each motion event, whether an utterance constitutes a motion event, which spatial concepts are expressed in the utterance, and which syntactic means are used to encode each concept. Figure 6 illustrates how shapes, colors, and symbols are combined to represent spatial meaning. In this way, the visual language of the application renders lexicalization patterns and preferences for verbalization directly observable.

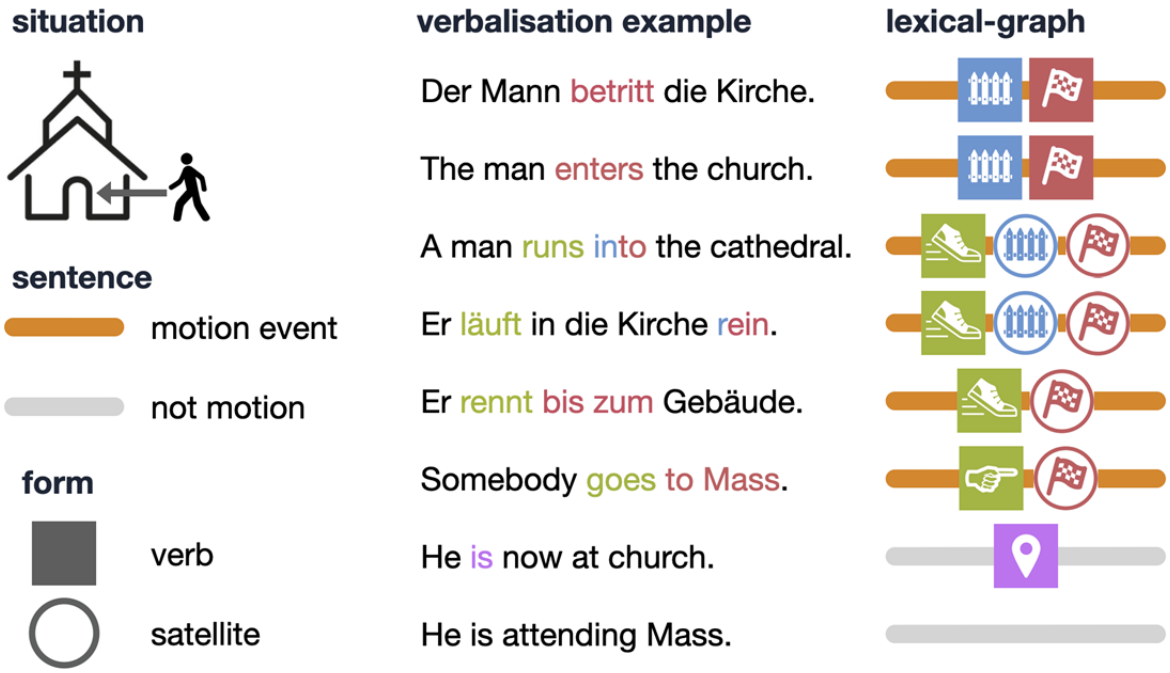

Figure 6: *Examples of the visual representation of the linguistic categories*

5. Class design and schedule

The primary objective of the course was to enable students to discover language-specific preferences in their own first language by exploring a dataset of motion event verbalizations. The course was conducted across four cohorts in consecutive semesters over a two-year period and involved a total of 67 students. The first cohort took place in the winter semester 2021/22 and consisted of 24 bachelor's students. The second cohort was held in the summer semester 2022 with eight master's students. The third cohort took place in the winter semester 2022/23 and included 19 bachelor's students, while the fourth cohort ran in the summer semester 2023 with 16 bachelor's students. Across all cohorts, classes consisted primarily of students of German language and literature, alongside a smaller number of students specializing in English language studies. All participants were prospective language teachers enrolled at a German university. A central element of the course involved working with linguistic data collected from individuals personally known to the students, thereby fostering engagement and a sense of ownership over the material. The semester-long course (15 sessions of 90 minutes each) began with a discussion of what qualifies a speaker as native or advanced (Class 1).

In the following week (Class 2), each student invited two participants from their social network to take part in the study and collected oral data using an abbreviated version of the classical motion event elicitation task (von Stutterheim et al. 2012). Participants included one functionally monolingual German speaker and one highly proficient German speaker of English as a foreign language. Functionally monolingual participants were typically older relatives, whereas advanced English speakers were younger professionals working in international contexts or fellow students of English studies. The English-speaking group was therefore, on average, younger and had a higher level of educational attainment, based on self-report. Across all participants, approximately 70% identified as female and 30% as male. Classes 3 and 4 focused on the German motion verb lexicon, examining intransitive motion verbs and discussing semantic nuances in verbs such as *gehen, laufen, flitzen, kraxeln, rennen, robben, betreten* and *überqueren*. When prompted, students predominantly listed manner-of-motion verbs, reflecting the prototypical status of manner encoding in German. Classes 5 and 6 continued with guided brainstorming on lexical and grammatical contrasts between English and German. Drawing on their own experiences as learners and prospective teachers, students identified features of English that are particularly challenging for German speakers. In parallel with these sessions, the instructor (acting as the comparative linguist) transcribed, segmented, coded, and entered the collected data into the application.

After coding was completed, Class 7 introduced the *LexiGraph* application and its interface. Students were informed that the app visually represented the semantic and syntactic information contained in the collected verbalizations, but the specific meanings of colors, shapes, and symbols were not yet disclosed. They were encouraged to explore the interface freely in order to become familiar with its underlying logic. Classes 8 and 9 were devoted to learning how to read and interpret the visual code. Each student received a booklet translating their own collected verbalizations into the visual format: a "Rosetta Stone" for decoding the app (see Figure 7). Working in pairs, students examined the visualized verbalizations of their own participants and attempted to infer the underlying coding system.

Figure 7: *Example of an English-language booklet used to support decoding of the visual coding system*

In Class 10, students compared and discussed their hypotheses about the meaning of the symbols. The instructor revealed the correct interpretations only when no group independently arrived at them. Classes 11 to 13 were exploratory sessions in which students used the application to identify cross-linguistic patterns in the dataset. Figure 8 shows the app in use. Students worked in small groups on guided discovery tasks, such as identifying the most frequent verbalizations for individual videos, locating speakers who used consistent verbalization schemas across explicit goals, examining the conditions under which sources were mentioned, or determining which meanings German verbs encoded most frequently. The aim of these tasks was for students to formulate hypotheses about differences between English and German in the expression of motion events.

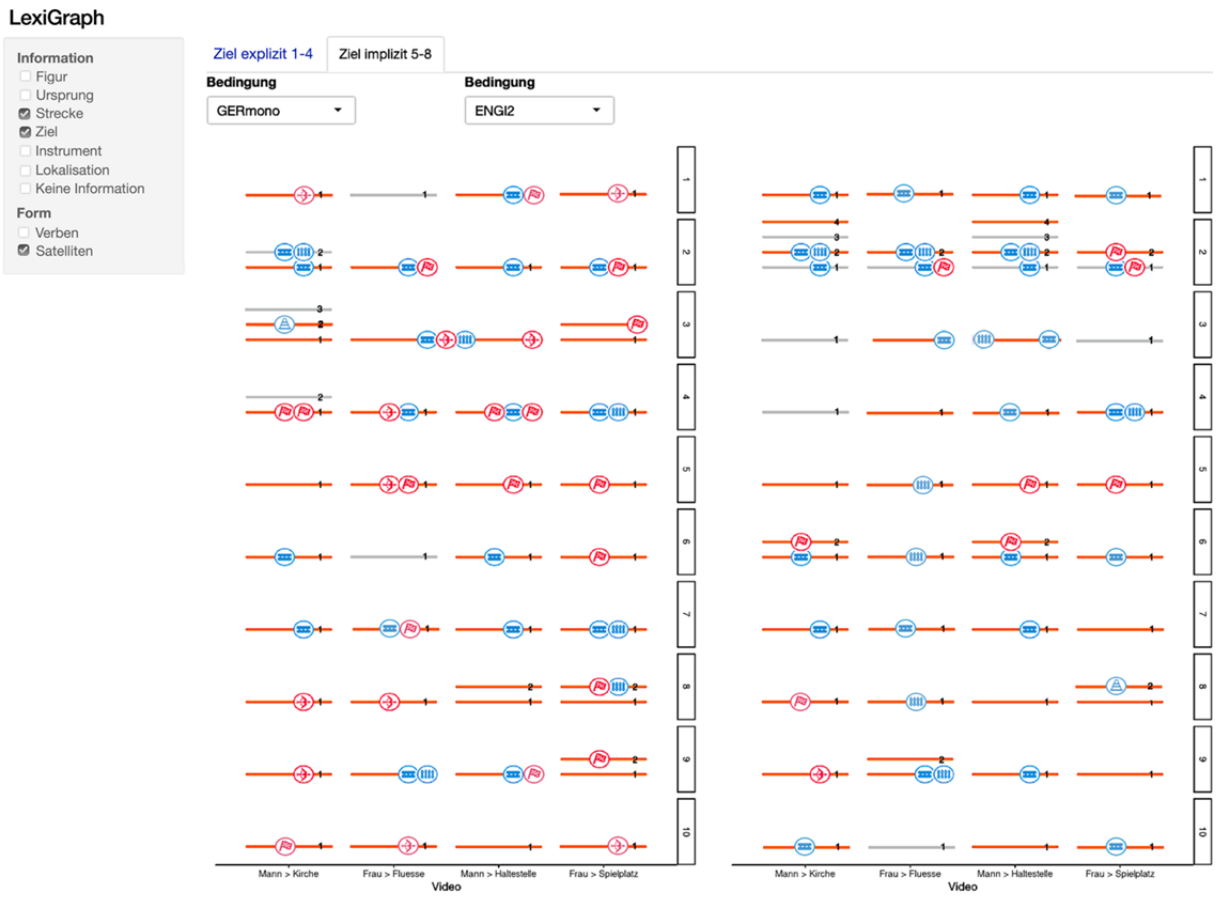

Figure 8: *Example of LexiGraph in use, displaying satellites encoding trajectory and goal*

Class 14 was devoted to a plenary discussion in which students' findings were compared with established literature on motion event typology. Drawing on assigned readings, students examined whether patterns reported in the literature were reflected in their own corpus, with particular attention to goal salience and lexicalization patterns. Homework included a short questionnaire evaluating the course. In the final session, students completed a written test assessing both theoretical understanding of cross-linguistic preferences and the ability to apply this knowledge.

The pedagogical design of the course relies on independent, data-driven exploration supported by indirect instructor guidance. The combination of practical (hands-on) and

theoretical (top-down) components encourages students to approach grammar inductively (Widodo 2006). A key advantage of using the application, rather than relying solely on direct instruction, lies in its experiential and discovery-based nature. Traditional grammar teaching presents linguistic rules declaratively, whereas this approach replicates the process of uncovering patterns from input, which is the same challenge faced by actual language learners. By exploring authentic data visually, students hypothesize, test, and refine their interpretations, fostering deeper cognitive engagement and retention (Meyer & Land 2005). The visual medium also renders abstract semantic-syntactic relations tangible and comparable across speakers and situations. This experience allows prospective teachers not only to grasp the theory of language-specific preferences, but also to experience the difficulty of deriving such knowledge from input alone. Explicit instruction rarely provides such an awareness. To date, the application and accompanying course design have been implemented across four student cohorts at both bachelor's and master's levels. Feedback was predominantly positive: students described the app as creative and engaging, although some perceived it as overly complex, too abstract, or only indirectly related to their future teaching practice. Overall, about 94% of students passed the course, and about 63% indicated that they would recommend it to others.

6. Planned improvements to the application

This project illustrates how innovative web applications can support both teachers and students in engaging with the complexity of linguistic data. At the present stage, the primary goal was to develop and test a minimum viable product, demonstrating that the application functions as intended and can be effectively integrated into teacher training contexts. Having achieved this goal, the next phase of development focuses on assessing the sustainability and scalability of the approach.

The initial development phase required substantial effort to design the application and prepare the underlying database. Now that a functional prototype exists, future work can prioritize efficiency and extensibility rather than fundamental design changes. One planned enhancement is the addition of hover text displaying the original verbalization in textual form when users interact with the visual symbols. A central focus of future development concerns the partial automation of the annotation process, particularly the coding of linguistic data. The proposed system will be based on a lightweight, modular workflow drawing on Universal Dependencies (UD; Schacht & Delucchi 2025). Using established UD parsers such as Stanza or UDPipe, transcribed verbalizations will be automatically tokenized, lemmatized, and syntactically parsed in order to identify relevant grammatical relations (e.g., *obl, nmod, case*) and morphological features (e.g., *Case, upos, deprel*). Within this parsed representation, prepositional phrases associated with motion verbs will be extracted and categorized using a combination of rule-based and machine-learning approaches. Prepositions and case markers will function as high-precision cues, e.g., *aus* or *von* for SOURCE, *in* + accusative, *nach*, or *zu* for GOAL, and *über* or *durch* for TRAJECTORY. To improve recall and disambiguation, a simple supervised classifier (e.g., a BiLSTM-CRF model with *fastText* embeddings) will

complement the rule-based component by assigning semantic roles to ambiguous prepositional phrases. The resulting annotations will then be integrated directly into *LexiGraph*'s visualization layer, enabling automatically annotated motion structures to be displayed in the same format as the currently manually coded data.

Once implemented, these enhancements are expected to substantially reduce annotation time, increase reproducibility, and facilitate the analysis of larger datasets. The use of the same UD-based feature set as in the ExpLay pipeline (Schacht & Delucchi 2025) ensures methodological continuity and compatibility with future extensions. For quality assurance, a linguist will continue to conduct targeted spot checks to verify the accuracy of the automatic annotations before datasets are made available to students.

Overall, these planned improvements will increase the scalability and long-term sustainability of the application for both teaching and research purposes, while preserving its pedagogical focus on discovering language-specific patterns through data exploration. Although the current version of *LexiGraph* is based on a single dataset, its architecture is readily extensible to additional datasets of the same type (i.e., verbalizations of motion event videos), to other languages such as Spanish, Italian, or Turkish, and to different speaker groups (L1, L2, and foreign language learners).

One limitation of the current system is that the visualization framework was specifically designed to represent spatial information within the semantic domain of motion events. Extending the approach to other grammatical phenomena would therefore require the development of new visual symbols and interaction principles capable of representing different conceptual domains in an equally intuitive way.

7. Discussion

In the context of language teacher education at a German university, four cohorts of prospective teachers of German and English were trained to explore language-specific preferences in their first language, German, by examining a dataset of motion event verbalizations. This exploration was facilitated by an online application that employs a novel visual system to encode the semantic and syntactic information contained in the students' own data.

With prospective teachers and their future pupils in mind (including bilingual speakers and learners of German as a second language), the primary aim of this project was to sensitize teachers to the challenges involved in identifying language-specific preferences in another language. The application provides a distinctive opportunity for students to engage with authentic data produced by real speakers (often including themselves) and to uncover subtle regularities governing how meaning is mapped onto grammatical form across languages. Patterns described in the literature cannot always be readily identified within the app, reflecting the current state of motion event research, where findings are often probabilistic rather than categorical (Bepperling & Härtl 2013). Nevertheless, despite the initial unfamiliarity and complexity of the app's visual language, students were frequently able to identify recurring patterns independently.

The application supports the identification of well-established construction types such as satellite-framed and verb-framed constructions, as well as later extensions including equipollently framed and double-framing constructions. Identifying such construction types is challenging because their distribution depends on situational characteristics of the events being verbalized. For this reason, the app presents verbalizations grouped by situation type, enabling students to observe how linguistic choices vary systematically across contexts. In an initial iteration of the course, students were asked to annotate the data themselves. This approach was discontinued, as manual annotation proved too time-intensive and led to a high number of coding errors due to the students' limited experience. In the revised course design, annotation is therefore carried out by the linguist overseeing the class.

From an accessibility perspective, it must be noted that the app is not barrier-free, as it relies heavily on visual distinctions between symbols and colors. In addition, the method presupposes access to a computer with an internet connection, which may pose practical or financial constraints in some educational settings.

Pedagogically, the instructional approach aligns with principles associated with *data humanism* (Lupi 2017), which emphasize that narratives give meaning to data, while data lend credibility to narratives. The deliberately challenging interface of the application is intended to create a threshold learning experience that mirrors the cognitive complexity and occasional frustration of acquiring a new language. The use of students' own verbalization data further personalizes the learning experience and increases engagement. While examining authentic data is widely assumed to be more convincing and conducive to long-term learning than theoretical discussion alone, this assumption remains an empirical question that should be addressed in future intervention studies comparing different instructional formats.

The application also facilitates discussion of key concepts such as second language users, foreign language learners, and bilingual speakers. Importantly, working with authentic data challenges the notion of monolingual speakers as "ideal" or "perfect" users of a language by revealing variation even among native speakers.

Reflecting on our experience, we recommend that language teacher programs consider incorporating foreign language training in languages typologically distant from the prospective teachers' first language. Learning typologically close languages often does not create sufficient distance from one's native linguistic system; as a result, relatively high proficiency can be achieved by simply "translating" from the L1 into the target language, without developing language-specific patterns of *thinking for speaking* in the other language. In contrast, engagement with a language that differs substantially in lexicon, syntax, or writing system can generate an experience of linguistic foreignness which language learners do not experience with languages closer to their L1. For prospective teachers, the primary benefit of learning a foreign language may lie in the cognitive and linguistic experience of stepping outside their comfort zone and no longer being able to rely on the grammar and vocabulary of their first language. This experience can be cognitively and emotionally demanding, often involving feelings of uncertainty, inadequacy, or failure. Such negatively charged experiences, however, are central to threshold learning, as they prompt learners to abandon familiar assumptions and adopt new perspectives. Experiencing this process may later enable teachers to empathize more deeply with pupils who struggle to acquire the languages they teach.

For teachers who have already mastered the languages they teach, working with *LexiGraph* can serve as a proxy for learning a typologically distant language. The app's unfamiliar visual language forces users to move beyond habitual modes of linguistic reasoning, creating a form of cognitive dissonance comparable to that experienced by language learners. This process not only raises awareness of cross-linguistic variation but also simulates the discomfort of being unable to rely on native-language intuition, which is a prerequisite for genuine threshold learning.

In this way, *LexiGraph* engages students in a form of threshold learning that extends beyond intellectual understanding. By grappling with an unfamiliar visual system and uncovering latent linguistic patterns, prospective teachers experience (on a limited scale) the cognitive and emotional challenges faced by language learners. This encounter fosters empathy toward learners who must infer grammatical principles from input alone. The resulting awareness is not only metalinguistic, but also ethical, as it cultivates sensitivity to the cognitive effort and vulnerability inherent in multilingual education. Through the integration of theory, empiricism, and experiential learning, *LexiGraph* contributes to inclusive teacher training by transforming abstract knowledge about language-specific preferences into embodied understanding, preparing future educators to support linguistically diverse learners with greater insight and compassion.

References

Alghamdi, A., M. Daller, and J. Milton. 2019. The persistence of L1 patterns in SLA: The boundary crossing constraint and incidental learning. *Vigo International Journal of Applied Linguistics* 16: 81–106.

Antonijević, S., and S. Berthaud. 2009. Verbs of motion and sentence production in second language. In *Proceedings of the 25th Annual Meeting of the International Society for Psychophysics (Fechner Day)*, Galway, Ireland, 487–492.

Aske, J. 1989. Path predicates in English and Spanish: A closer look. In K. Hall, M. Meacham, and R. Shapiro (eds.), *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society*, 1–14. Berkeley: Berkeley Linguistics Society.

Athanasopoulos, P., and E. Bylund. 2013. Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science* 37: 286–309.

Athanasopoulos, P., L. Damjanovic, J. Burnand, and E. Bylund. 2015. Learning to think in a second language: Effects of proficiency and length of exposure in English learners of German. *Modern Language Journal* 99(1): 138–153.

Attwood, A. 2014. The teachability of entering and exiting in L2 English. *Language Studies Working Papers* 6: 43–52. University of Reading.

Beare, K. 2015. Visual explanations of each English tense. Available online at https://www.thoughtco.com/english-grammar-tenses-chart-4123178, accessed 22 December 2025.

Beavers, J., B. Levin, and S. W. Tham. 2010. The typology of motion expressions revisited. *Journal of Linguistics* 46(2): 331–377.

Bepperling, S., and H. Härtl. 2013. Ereigniskonzeptualisierung im Zweitspracherwerb – Thinking for speaking im Vergleich von Muttersprachlern und Lernern. *Zeitschrift für Semiotik* 35(1–2): 159–191.

Berman, R. A., and D. Slobin (eds.). 1994. *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Erlbaum.

Bohnemeyer, J., N. J. Enfield, J. Essegbey, I. Ibarretxe-Antuñano, S. Kita, F. Lüpke, and F. K. Ameka. 2007. Principles of event segmentation: The case of motion events. *Language* 83: 495–532.

Bylund, E. 2008. Conceptualization processes of events in Spanish and Swedish. *Revue Romane* 43(1): 1–24.

Cadierno, T. 2004. Expressing motion events in a second language: A cognitive typological perspective. In M. Achard and S. Niemeier (eds.), *Cognitive linguistics, second language acquisition and foreign language teaching*, 13–49. Berlin: Mouton de Gruyter.

Cadierno, T. 2008. Motion events in Danish and Spanish: A focus-on-form pedagogical approach. In S. de Knop, T. de Rycker, and R. Dirven (eds.), *Cognitive approaches to pedagogical grammar: A volume in honour of René Dirven*, 259–294. Amsterdam: John Benjamins.

Cadierno, T., and P. Robinson. 2009. Language typology, task complexity and the development of L2 lexicalization patterns for describing motion events. *Annual Review of Cognitive Linguistics* 7: 245–276.

Cadierno, T., and L. Ruiz. 2006. Motion events in Spanish L2 acquisition. *Annual Review of Cognitive Linguistics* 4(1): 183–216.

Carroll, M., C. von Stutterheim, and R. Nüse. 2004. The language and thought debate: A psycholinguistic approach. In C. Habel and T. Pechmann (eds.), *Approaches to language production*, 183–218. Berlin: de Gruyter.

Chen, L., and J. Guo. 2009. Motion events in Chinese novels: Evidence for an equipollently-framed language. *Journal of Pragmatics* 41(9): 1749–1766.

Croft, W. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Croft, W. 2003. *Typology and universals*. Cambridge: Cambridge University Press.

Croft, W., J. Barðdal, W. Hollmann, V. Sotirova, and C. Taoka. 2010. Revising Talmy's typological classification of complex event constructions. In H. C. Boas (ed.), *Contrastive studies in construction grammar*, 201–235. Amsterdam: John Benjamins.

Delucchi Danhier, R. 2017. *Weganweisungen im Deutschen und Spanischen: Eine Vergleichsanalyse unter Anwendung von Visualisierungen*. München: Verlag Dr. Hut.

Delucchi Danhier, R., and P. González Ávalos. LexiGraph: A web application for visualizing motion event lexicalization patterns. Available online at https://pga99.shinyapps.io/VSM_Vis/, accessed 22 December 2025.

Filipović, L. 2007. *Talking about motion: A cross-linguistic investigation of lexicalization patterns*. Amsterdam: John Benjamins.

Georgakopoulos, T., H. Härtl, and A. Sioupi. 2019. Goal realization: An empirically based comparison between English, German and Greek. *Languages in Contrast* 19(2): 280–309.

Hill, D. J. 1992. Cluster analysis and the interlanguage lexicon. *Edinburgh Working Papers in Applied Linguistics* 3: 67–77.

Johanson, M., and A. Papafragou. 2010. Universality and language specificity in the acquisition of path vocabulary. In K. Franich, K. M. Iserman, and L. L. Keil (eds.), *Proceedings of the 34th Annual Boston University Conference on Language Development*, 185–196. Somerville, MA: Cascadilla Press.

Lakoff, G. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Laws, J., A. Attwood, and J. Treffers-Daller. 2021. Unlearning the boundary-crossing constraint: Processing instruction and the acquisition of motion event construal. *International Review of Applied Linguistics in Language Teaching* 60(4): 1089–1118.

Lupi, G. 2017. Data humanism: The revolutionary future of data visualization. *PrintMag*. Available online at https://www.printmag.com/article/data-humanism-future-of-data-visualization/, accessed 22 December 2025.

Meyer, J., and R. Land. 2005. Threshold concepts and troublesome knowledge (2): Epistemological considerations and a conceptual framework for teaching and learning. *Higher Education* 49: 373–388.

Montero-Melis, G., S. Eisenbeiss, B. Narasimhan, I. Ibarretxe-Antuñano, S. Kita, A. Kopecka, F. Lüpke, T. Nikitina, I. Tragel, T. F. Jaeger, and J. Bohnemeyer. 2017. Satellite- vs. verb-framing underpredicts nonverbal motion categorization: Insights from a large language sample and simulations. *Cognitive Semantics* 3: 36–61.

Mora Gutiérrez, J. P., and P. Antela Costa. 2021. Teaching the difference between English and Spanish with regard to the expression of directed motion using a gamified conceptual workshop. Paper presented at the

workshop *Can Motion Event Construal be Taught or Reconstructed? Evidence from Bilinguals and L2 Learners*, University of Reading, April 2021.

Navarro, S., and E. Nicoladis. 2005. Describing motion events in adult L2 Spanish narratives. In D. Eddington (ed.), *Selected proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, 102–107. Somerville, MA: Cascadilla Press.

Park, H., S. Choi, and M. Lee. 2012. Visual input enhancement, attention, grammar learning, and reading comprehension: An eye movement study. *English Teaching* 67(4): 241–265.

R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available online at https://www.R-project.org/, accessed 22 December 2025.

Schacht, C., and R. Delucchi Danhier. 2025. ExpLay: A new corpus resource for research on expertise as an influential factor on language production. In *Proceedings of the 19th Linguistic Annotation Workshop (LAW-XIX)*, 216–227. Vienna: Association for Computational Linguistics.

Schlyter, S. 1984. L'acquisition des verbes de déplacement/mouvement par les adultes suédois. *Papers from the Institute of Linguistics, University of Stockholm* 52.

Schlyter, S., and Å. Viberg. 1985. French and Swedish as targets in second language lexical acquisition. *Scandinavian Working Papers in Bilingualism* 4: 110–135.

Schmiedtová, B. 2011. Do L2 speakers think in the L1 when speaking in the L2? *International Journal of Applied Linguistics* 8: 138–179.

Schmiedtová, B., and M. Flecken. 2008. Aspectual concepts across languages: Some considerations for second language learning. In S. de Knop and T. de Rycker (eds.), *Cognitive approaches to pedagogical grammar*, 357–384. Berlin: Mouton de Gruyter.

Slobin, D. 1991. Learning to think for speaking: Native language, cognition, and rhetorical style. *Pragmatics* 1(1): 7–25.

Slobin, D. 1996a. From "thought and language" to "thinking for speaking." In J. J. Gumperz and S. C. Levinson (eds.), *Rethinking linguistic relativity*, 70–96. Cambridge: Cambridge University Press.

Slobin, D. 1996b. Two ways to travel: Verbs of motion in English and Spanish. In S. Thompson and M. Shibatani (eds.), *Grammatical constructions: Their form and meaning*, 195–219. Oxford: Clarendon Press.

Slobin, D. 2004. The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömqvist and L. Verhoeven (eds.), *Relating events in narrative: Typological and contextual perspectives*, 219–257. Mahwah, NJ: Erlbaum.

Talmy, L. 1972. *Semantic structures in English and Atsugewi*. Ph.D. diss., University of California, Berkeley.

Talmy, L. 2000. *Toward a cognitive semantics*. Cambridge, MA: MIT Press.

Talmy, L. 2023. Foreword: A taxonomy of cognitive semantics. In F. T. Li (ed.), *Handbook of cognitive semantics*. Leiden: Brill.

Treffers-Daller, J., and F. Tidball. 2015. Can L2 learners learn new ways to conceptualise events? Evidence from motion event construal among English-speaking learners of French. In P. Guijarro-Fuentes, K. Schmitz, and N. Müller (eds.), *The acquisition of French in multilingual contexts*, 145–184. Bristol: Multilingual Matters.

von Stutterheim, C., and R. Nüse. 2003. Processes of conceptualisation in language production. *Linguistics* 41(5): 851–881.

von Stutterheim, C., M. Andermann, M. Carroll, M. Flecken, and B. Schmiedtová. 2012. How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye-tracking data, and memory performance. *Linguistics* 50(4): 833–867.

Widodo, H. P. 2006. Approaches and procedures for teaching grammar. *English Teaching: Practice and Critique* 5(1): 122–141.

Cite this article as:

Delucchi Danhier, R., & González Ávalos, P. (2025). Teaching Language-Specific Preferences in the Native Language: An Interactive Web App for Prospective Teachers. *LingBaW. Linguistics Beyond and Within, 11*, 69–88.