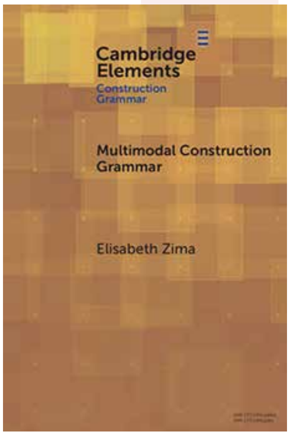
# **Beyond words:** Building a multimodal construction grammar

BY: **Georgios Damaskinidis**



Elisabeth Zima
**Multimodal Construction Grammar**

Cambridge: Cambridge University Press, 2025, pp. 62,
ISBN: 9781009359856 (ebook), ISBN: 9781009359870 (Pbk).

ARTICLE INFO:
Volume: **12**
Issue: **02**
Winter **2025**
ISSN: **2459-2943**
DOI: **10.18680/hss.2025.0028**
Pages: **177-184**
Lic.: **CC BY-NC-ND 4.0**

Elisabeth Zima's *Multimodal Construction Grammar* arrives precisely when construction-based theories of grammar are addressing the fact that everyday language use is audiovisual, embodied, and highly interactional. Instead of viewing gesture, gaze, prosody, and posture as mere 'add-ons' to verbal structure, the book argues that a credible construction must represent patterned combinations across modalities. Framed firmly within usage-based Construction Grammar and continuously engaging with Conversation Analysis and interactional linguistics, Zima provides both a comprehensive overview of recent findings and a thoughtful proposal for how we might model them.

Zima argues that multimodality is not a peripheral curiosity but a constitutive dimension of many constructions. That claim is intentionally moderate: the book does not suggest that *all* constructions are inherently multimodal; instead, it shows that a nontrivial subset displays entrenched, functionally relevant cross-modal patterns. The Element is written for three overlapping readerships. First are construction grammarians who need a principled way to incorporate bodily conduct and prosody into their representational machinery. Second are interactional linguists and conversation analysts, whose detailed descriptions of 'multimodal packages' and 'assemblies' (e.g., Bressem and Müller 2017; Stukenbrock 2021) provide rich empirical grist but often stop short of cognitive commitments. Third, are corpus and tool developers building the archives and annotation pipelines that enable quantitative generalizations (see also Zima 2020).

The core thesis is straightforward: if constructions are conventionalized form-meaning pairings learned from use, and if the evidence of use is routinely audiovisual, then a constructicon that ignores nonverbal form underestimates what is stored, learned, and processed. The challenge, as Ziem (2017) pointedly asked, is whether we 'really need' a multimodal construction grammar or whether looser notions of accompaniment suffice. Zima's answer is empirical and constructive: sometimes an association is enough; sometimes the nonverbal layer is integral and therefore part of the construction's form.

Historically, the 'multimodal turn' has roots in gesture studies, prosody in interaction, and ethnomethodological CA (e.g., Ogden 2010; Ward 2019; Stukenbrock 2010). Zima synthesizes this literature with constructionist commitments about entrenchment and schematicity. She sketches how early demonstrations of gesture-speech coordination (e.g., deictic and depictive gestures that make spatial dimensions available to recipients) pushed linguistics beyond transcripts of words alone. Subsequent work identified *recurrent* cross-modal patterns that behave like construction families (Bressem and Müller 2017; Ningelgen and Auer 2017), strengthening the case that multimodal phenomena are not merely local conveniences but patterned resources with learnable form – function pairings. In parallel, constructionist studies began to quantify gesture–speech coupling for specific constructions (Zima 2014, 2017a, 2017b), while interactional research charted how gaze and body orientation scaffold sequential organization (Zima 2020; Stukenbrock 2021). The Element knits these strands together into a coherent research program.

One of the book's most useful contributions is a clear typology of representational options. Zima reconstructs three families of proposals, each motivated by different data segments.

1. Obligatoriness. In a stringent view, a multimodal construction exists when the nonverbal component is *required* for interpretation. Deictic and depictive patterns with German so – for instance, *so groß* 'this big' accompanied by a size-depicting gesture – are exemplary: without the gesture, the intended meaning is underspecified (Stukenbrock 2010; Ningelgen & Auer 2017). This makes the case for representations that *include* gesture as part of constructional form. At the same time, as Ziem (2017) cautions, 'obligatoriness' sets a very high bar, and only a small corner of multimodal phenomena will qualify.

2. Cross-modal association. A more graded alternative treats many multimodal patterns as robust associations between otherwise unimodal constructions and recurring nonverbal behaviors. Uhrig (2022), for example, uses collostructional methods to show how hand-gesture families pattern with English verbs of throwing, revealing meaning-sensitive tendencies without making them obligatory. Large-scale, distributional approaches to time expressions that integrate gesture (Pagán-Cánovas et al. 2020) converge in showing systematic skew in the gesture–speech relationship. Zima favors modeling such associations in a network with weighted links, acknowledging gradience and variability.
3. Utterance Construction Grammar (UCxG). Cienki (2017) proposes that constructions are licensed at the level of *utterance structure*, with deep-structure representations encoding multimodal potential and surface-structure realizations instantiating variable subsets (verbal, gestural, prosodic). Zima treats UCxG as an attractive compromise: it respects the event-based nature of interaction (a CA insight), offers representational slots for nonverbal form, and avoids pure associationism that merely lists co-occurrences.

Zima resists the urge to declare a winner. Instead, she reframes the problem as one of prediction: which representational scheme best accounts for distributional facts, psycholinguistic discriminations, and learnability constraints? That orientation is particularly welcome in a literature that can tilt theoretical.

A strength of the Element is its methodological literacy. Zima compiles quantitative results on gesture–speech co-occurrence across construction types, illustrating that coupling varies by constructional family: motion and distance constructions often recruit depictive gesture; stance-laden idioms can have distinctive tempo and intonation profiles; modal particles show weaker association with gesture. Rather than canonizing raw proportions, the book highlights tools that quantify association with effect sizes and control for dispersion, semantic class, and register (Uhrig 2022; Pagán-Cánovas et al. 2020). Zima (2017a, 2017b) shows, for instance, how the English [*all the way from X PREP Y*] construction exhibits recurrent gesture types that cluster semantically; Hinell (2018) documents aspectual auxiliaries' gestural profiles; and Bressem and Müller (2017) identify a recurrent 'throwing-away' gesture that functions as part of a broader negative-assessment resource.

Prosody receives especially careful treatment. Building on Ward's (2019) inventory of English prosodic patterns and Ogden's (2010) work on prosody in complaints, Zima reviews recent experiments demonstrating that listeners discriminate constructional senses using prosody alone. Lehmann's (2024a) study of *Tell me about it* shows that the stance use and the information-request use differ in tempo and that naive listeners reliably sort tokens by use based on audio stripped of lexical content; Lehmann (2024b) further develops the idea of a prosodic 'mode' as a meaning-bearing, constructional dimension in its own right. Italian 'list constructions' offer another window: Masini, Combei, and Cicchirillo (in press) show how articulation rate and tonal parallelism support list interpretation, reinforcing the claim that prosodic features are part of stored constructional knowledge. Zima weaves these findings into a persuasive case for treating prosodic structure as a constructional form rather than merely as performance.

Gaze and recipient design, though less frequently operationalized in CxG, are also foregrounded. Zima (2020) demonstrates how gaze behavior in triadic storytelling aligns with turn organization and recipient feedback, underscoring that a multimodal constructicon must accommodate gaze as both a resource for action and a recurrent formal pattern. That perspective dovetails with Stukenbrock's (2021) notion of 'multimodal gestalts,' whose routinization may resemble grammaticalization over time.

The book's dialogue with CA is particularly fruitful. CA research prioritizes sequential organization and often deliberately brackets cognitive representation; yet it routinely documents recurring couplings of linguistic formats with bodily conduct that accomplish recognizable social actions. Bressem and Müller's (2017) 'negative-assessment construction,' built around a recurrent throwing-away gesture, is a case in point. Zima argues that such findings are ideal starting points for constructional analysis: they identify candidate patterns, delimit their interactional ecology (e.g., dispreferred turn shapes, assessments, stance displays), and suggest functional generalizations. The challenge, one the Element takes up without polemic, is to move from 'recurrent multimodal package' to 'stored construction' responsibly, with converging evidence from distributional tendencies, processing, and learning.

German *so* again serves as a touchstone. Multiple subpatterns – *so groß* with size-depicting gesture; *so machen* with iconic enactment; *so sieht/sehen X* aus with visible presentation – appear to *require* gesture for full interpretability (Stukenbrock 2010; Ningelgen and Auer 2017). Those are strong candidates for obligatoriness. Elsewhere, the pairings are looser: negative assessments can be realized with or without the throwing-away gesture; stance idioms like *Tell me about it* are disambiguated by prosody but still comprehensible without it (Bressem and Müller 2017; Lehmann 2024a). Zima's point is that a single, monolithic criterion ('obligatory or not?') is ill-suited to this landscape; graded, networked representations are better.

If we grant that some degree of multimodal coupling is entrenched and meaningful, how should a usage-based grammar store it? Zima surveys three representational strategies: (i) enlarging constructional *form* to include prosodic, gestural, and gaze features (Ward 2019; Masini et al. in press); (ii) positing *associative links* between verbal constructions and nonverbal form schemas with weights reflecting strength (Uhrig 2022); and (iii) adopting utterance-level deep/surface structure that enumerates modality-specific slots (Cienki 2017). Each choice brings trade-offs in redundancy, learnability, and predictive power. Ziem’s (2017) skepticism usefully keeps the bar high: if every co-occurrence is stored, the constructicon risks becoming an unprincipled scrapbook. Zima meets that worry by urging explicit thresholds derived from effect sizes, dispersion measures, and functional specificity, an agenda that points directly to corpus and experimental tests.

Zima repeatedly emphasizes that progress depends on the scale and quality of the data. Many tantalizing patterns sit in the medium-frequency range, invisible to small corpora and fragile under naive counting. The book calls for larger, higher-fidelity, and better-synchronized audiovisual archives; for improved annotation of gesture phases and alignment windows; and for tools that make model-based inference accessible to linguists. The ‘discovery vs. test’ cycle she proposes is at once modest and ambitious: use CA-inspired close analysis to uncover candidate patterns (Bressem and Müller 2017; Ogden 2010), quantify their distribution with collostructional or distributional methods (Uhrig 2022; Pagán-Cánovas et al. 2020), and then test processing consequences (Lehmann 2024a, 2024b). Zima’s (2014, 2017a) own case studies on English motion/distance constructions exemplify this pipeline, as do her and others’ calls to include gaze via mobile eye-tracking in naturalistic settings (Zima 2020).

A major virtue of the Element is its even-handed conceptual framing. Zima neither collapses all multimodal phenomena into “constructions” nor sets an impossibly strong criterion that would make the category nearly empty. By juxtaposing obligatoriness (as in deictic *so*: Stukenbrock 2010; Ningelgen and Auer 2017), strong but non-obligatory association (as in throwing verbs and recurrent gesture families: Uhrig 2022), and utterance-level licensing (Cienki 2017), she clarifies what is at stake theoretically and points to the kinds of evidence that should decide it.

The book treads nimbly between close, sequential analysis and large-scale quantitative work. It takes CA’s descriptive discipline seriously, treating interactional ‘packages’ as indispensable for discovering patterns (Bressem & Müller 2017; Ogden 2010), but it is equally clear that robust generalization demands numbers, ideally, numbers that respect dispersion, register, and semantic class (Uhrig 2022; Pagán-Cánovas et al. 2020). The prosody chapters are exemplary here: Ward (2019) and Ogden (2010) provide inventories and interactional settings; Lehmann (2024a, 2024b) adds processing evidence; Masini et al. (in press) offer construction-specific prosodic profiles. Zima doesn’t oversell any single method; she shows how they fit together.

A persistent bias in multimodal work is the equating of ‘multimodality’ with manual gesture. Zima corrects that bias by according equal status to prosody and gaze. The cumulative case for ‘prosodic constructions’ now seems compelling, from English stance expressions to list structures (Ward 2019; Lehmann 2024a; Masini et al., in press). Gaze is admittedly more difficult to capture and operationalize, but Zima (2020) demonstrates its patterned role in turn organization, and Stukenbrock (2021) shows how routinized visual gestalts can stabilize into recognizable formats. For a constructicon that claims to mirror usage, excluding these channels would be an empirical mistake.

The Element's tone is integrative rather than polemical. It takes Ziem's (2017) question about the necessity of multimodal CxG seriously, answers it by showing cases where a constructional analysis yields explanatory power, and remains candid about where the evidence is not yet decisive. By inviting CA researchers to consider representational payoffs and CxG scholars to reckon with sequential organization, Zima models the kind of cross-tradition dialogue the field needs.

First, readers may want more operational guidance on *thresholds*, how strong, how dispersed, and how functionally specific a cross-modal association should be before we posit a stored construction. Zima gestures toward decision criteria, but a worked-out decision tree that combines effect size, dispersion, and functional profiling would accelerate uptake. Second, the empirical base remains richer for English and German than for other languages. Cross-linguistic work on deictic-gesture obligatoriness, stance-prosody coupling, and list prosody would sharpen theoretical claims (see Zima 2017b; Stukenbrock 2021, for initial steps). A third desideratum is a more explicit representational format for prosody: Ward (2019) provides a descriptive inventory, and Lehmann (2024b) argues for prosodic 'modes,' but a sketch of how, say, an Italian list construction would look in a prosody-rich constructicon would be instructive (Masini et al., in press).

Finally, the acquisition-and-processing story is still developing. Lehmann's (2024a) forced-choice and discrimination studies are exemplary; analogous experiments on gesture-speech coupling and gaze-conditioned interpretations would round out the picture. How children acquire these cross-modal pairings – what generalizations they extract, how they weigh channels, and how quickly they entrain to community-specific gestural conventions – remains a compelling open agenda (see Pagán-Cánovas et al. 2020, for large-scale evidence relevant to learning).

Although the Element is tightly argued, it is not forbiddingly technical. Zima writes with an expository clarity that will help readers from different traditions meet on common ground. Summaries of key studies are concise and precise – e.g., Zima's (2017a, 2017b) own work on English circular motion and distance constructions; Pagán-Cánovas et al.'s (2020) dataset-driven approach to time expressions; Hinell's (2018) analysis of aspectual auxiliaries; and Lanwer's (2017, 2020) work on apposition and prosody – which makes the book suitable for graduate seminars. The running dialogue with skeptics (Ziem 2017) and bridge builders (Cienki 2017) keeps the text honest about what is known and what remains unsettled.

*Multimodal Construction Grammar* is both a lucid primer and a research agenda. It shows where the strongest evidence lies (obligatory deictic-gesture couplings; robust, graded gesture–speech associations; prosodic constructions), where modelling choices matter (obligatoriness vs. association vs. utterance-level structure), and what it will take to make cumulative progress (bigger, better-annotated corpora; explicit thresholds; experimental tests) (Zima 2014, 2017a, 2017b, 2020, 2025). It bridges communities by showing how CA discoveries can seed constructional hypotheses and how constructionist representations can return explanatory dividends for interactional practice (Bressem and Müller 2017; Ogden 2010; Ward 2019). For anyone building a constructicon that aspires to reflect actual language use, Zima's Element belongs within arm's reach.

## References

Bressem, J., & Müller, C. 2017. The 'negative-assessment construction': A multimodal pattern based on a recurrent gesture? *Linguistics Vanguard 3*. https://doi.org/10.1515/lingvan-2016-0053

Cienki, A. 2017. Utterance Construction Grammar (UCxG) and the variable multimodality of constructions. *Linguistics Vanguard*, 3. https://doi.org/10.1515/lingvan-2016-0048

Debras, C. 2021. Multimodal profiles of *je (ne) sais pas* in spoken French. *Journal of Pragmatics* 182(1): 42–62.

Hinell, J. 2018. The multimodal marking of aspect: The case of five periphrastic auxiliary constructions in North American English. *Cognitive Linguistics* 29(4): 773–806.

Lanwer, J. 2017. Apposition: A multimodal construction? The multimodality of linguistic constructions in the light of usage-based theory. *Linguistics Vanguard* 3. https://doi.org/10.1515/lingvan-2016-0071

Lanwer, J. 2020. Appositive Syntax oder appositive Prosodie? In: W. Imo and J. Lanwer (eds.), *Prosodie und Konstruktionsgrammatik*. Berlin, Boston: De Gruyter, 233–281.

Lehmann, C. 2024a. Multimodal constructions revisited: Testing the strength of association between spoken and non-spoken features of *Tell me about it*. *Cognitive Linguistics* 35(3): 407–437.

Lehmann, C. 2024b. What makes a multimodal construction? Evidence for a prosodic mode in spoken English. *Frontiers in Communication 9*. https://doi.org/10.3389/fcomm.2024.1338844

Masini, F., Combei, C. R. and R. Cicchirillo (in press). The prosody of list constructions. In: K. Nikiforidou & M. Fried (eds.), *Multimodal communication from a construction grammar perspective*. Amsterdam: John Benjamins,116–151.

Ningelgen, J. and P. Auer 2017. Is there a multimodal construction based on non-deictic so in German? *Linguistics Vanguard 3*. https://doi.org/10.1515/lingvan-2016-0051

Ogden, R. 2010. Prosodic constructions in making complaints. In: D. Barth-Weingarten, E. Reber and M. Selting (eds.), *Prosody in interaction*. Amsterdam: John Benjamins, 81–104.

Pagán-Cánovas, C., Valenzuela, J., Alcaraz-Carrión, D., Olzá, I. and M. Ramscar 2020. Quantifying the speech–gesture relation with massive multimodal datasets: Informativity in time expressions. *PLOS ONE* 15(6), e0233892. https://doi.org/10.1371/journal.pone.0233892

Stukenbrock, A. 2010. Überlegungen zu einem multimodalen Verständnis der gesprochenen Sprache am Beispiel deiktischer Verwendungsweisen des Ausdrucks so. *InLiSt: Interaction and Linguistic Structures* 47: 1–23.

Stukenbrock, A. 2021. Multimodal gestalts and their change over time: Is routinization also grammaticalization? *Frontiers in Communication 6*, 662240. https://doi.org/10.3389/fcomm.2021.662240

Uhrig, P. 2022. Hand gestures with verbs of throwing: Collostructions, style and metaphor. In: B. Hampe and A. Binanzer (eds.), *Yearbook of the German Association of Cognitive Linguistics* (Vol. 10). Berlin, Boston: De Gruyter, 99–120.

Ward, N. G. 2019. *The prosodic patterns of English conversation*. Cambridge: Cambridge University Press.

Ziem, A. 2017. Do we really need a multimodal construction grammar? *Linguistics Vanguard* 3. https://doi.org/10.1515/lingvan-2016-0095

Zima, E. 2014. Gibt es multimodale Konstruktionen? Eine Studie zu [V(motion) in circles] und [all the way from X PREP Y]. *Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion* 15:1–48.

Zima, E. (2017a). On the multimodality of [all the way from X PREP Y]. *Linguistics Vanguard* 3. https://doi.org/10.1515/lingvan-2016-0055

Zima, E. 2017b. Multimodal constructional resemblance: The case of English circular motion constructions. In: F. Ruiz de Mendoza, A. Luzondo and P. Pérez-Sobrino (eds.), *Constructing families of constructions*. Amsterdam: John Benjamins, 301–337.

Zima, E. 2020. Gaze and recipient feedback in triadic storytelling activities. *Discourse Processes* 57(9): 725–748.

AUTHOR

**Georgios Damaskinidis**, Department of Psychology, University of Western Macedonia, Florina, Greece.