The L2 Network: A CEFR-Aligned Knowledge Graph for Grammar Domain Modeling

Luisa Ribeiro-Flucht, Xiaobin Chen
LEAD Graduate School and Research Network,
University of Tübingen, Germany
{luisa.ribeiro-flucht, xiaobin.chen}@uni-tuebingen.de

Abstract

Large language models have renewed interest in the role of structured linguistic data for applications that require controllable, interpretable, and pedagogically aligned language generation. This need is especially visible in intelligent language tutoring, where grammar cannot be modeled as a flat inventory of patterns alone, but must also capture their relations and functions they realize. We present the L2 Network, a machine-readable knowledge graph of CEFR A1-A2 English grammar that encodes formal patterns, functions, and typed relations between them. The resource is grounded in established pedagogical reference materials, combining form inventory and progression information from the English Grammar Profile with a functional layer derived from CEFR descriptors. We further report content validation of the form-function mappings through expert annotation, including agreement analysis and a consensus-filtered core release. The resulting graph provides an explicit schema for representing pedagogically relevant grammatical knowledge and supports downstream uses such as learner modeling, adaptive task selection, and controlled generation in dialogue-based ICALL systems.

Keywords: knowledge graph, structured linguistic data, CEFR, grammar modeling, ICALL

1. Introduction

Intelligent Computer-Assisted Language Learning (ICALL) has long sought to support second language (L2) learners through scalable and adaptive Intelligent Tutoring Systems (ITS). The recent rise of GenAI, particularly large language models (LLMs), has significantly expanded what is possible. Because LLMs can generate fluent, contextualized, and even multilingual text on demand, they have made it increasingly feasible to move beyond fixed exercise formats toward more open-ended, interactive forms of language practice, as well as to generate personalized content and individualized feedback. Recent work has therefore increasingly treated LLMs as a promising foundation for next-generation ITSs.

More fundamentally, however, this shift raises not only a tutoring challenge, but also a representational one. Language is increasingly understood as a multidimensional system in which formal patterns are associated with semantics, pragmatics, and discourse. For computational applications that require linguistic grounding and controlled generation, language knowledge therefore benefits from being represented as structured data. In ICALL, this need is especially visible: recent work shows that prompting LLMs to generate proficiency-aligned content or to act as adaptive tutors does not reliably produce pedagogically aligned output (Benedetto et al., 2025; Borchers and Shou, 2025; Jokinen, 2024).

Within ITS architectures, domain models provide an explicit representation of what is to be learned (Padayachee et al., 2002). In language tutoring, they have often taken the form of inventories of knowledge components or exercise-linked units, supporting exercise generation, feedback, and learner tracking (Katinskaia, 2025). While such representations can be adequate for relatively fixed exercise formats, they are more limited in open-ended interactive settings, where systems must do more than select target forms: they must also place them in meaningful contexts, guide generation toward relevant realizations, and interpret learner performance across related forms and functions. This motivates representing grammar as structured, relational data rather than as a flat inventory of items.

In this paper, we present the L2 Network as a structured linguistic resource and knowledge graph schema for representing CEFR-aligned English grammar in machine-readable form. The graph formalizes grammatical patterns (e.g., SUBJ have V-en), functions (e.g., describing experiences or expressing degree), and the typed relations that connect them, including realization, dependency, and progression.

Grounded in established pedagogical resources, the L2 Network is introduced both as an English-language resource and as a reusable modeling approach for transforming CEFR-aligned reference materials into structured, machine-readable domain models for other languages. Beyond its immediate value as an explicit domain representation, it also provides a foundation for future integration methods, such as graph-based retrieval, neuro-symbolic control, and graph-informed fine-tuning. The following sections describe its design principles, construction criteria, and initial validation.

Proceedings of the Workshop on Structured Linguistic Data and Evaluation, pages 148–159
May 11, 2026.

2. Background

To clarify what a pedagogically useful language domain model should represent, the present work draws on functional and cognitive traditions, which treat language as a system in which forms are tied to meaning and use (Behrens, 2009; Larsen-Freeman, 2003). Within this broader perspective, usage-based and constructionist approaches (Goldberg, 2003) analyze lexicon and grammar as a continuum of conventionalized form-meaning pairings, or constructions. These range from bound morphemes (e.g., third-person singular -s), to formulaic patterns (e.g., "would you mind ..."), to fully or partially schematic patterns (e.g., the X-er, the Y-er). On this view, linguistic knowledge is best represented as an interconnected network of constructions, a "constructicon" (Diessel, 2023). L2 development, in turn, can be understood as the gradual strengthening of mappings between learners' communicative intentions and the conventionalized forms of the target language that express them (Schmid, 2015).

These perspectives suggest that language is more appropriately modeled as a structured and interconnected system linking formal patterns to meanings and patterns of use, rather than as a flat inventory of isolated items. This raises a representational question: how can such multidimensional knowledge be formalized in a way that supports explicit encoding, interpretation, and computational use? In language education, one influential answer has been the functional-notional tradition, which organizes language around what learners are able to do with it in context rather than around forms alone. This orientation is carried forward in the Common European Framework of Reference for Languages (CEFR, (Council of Europe, 2001)), where proficiency is defined in terms of communicative competence.

In the CEFR, L2 competence is described through language-agnostic can-do statements that specify communicative achievements across proficiency levels. This reflects and extends the Council of Europe's earlier functional-notional tradition, especially Threshold Level 1990 (Van Ek et al., 1991), which organized language in terms of notions and communicative functions. For domain modeling, this is particularly valuable because it provides a principled functional layer: an abstract inventory of communicative goals that can serve as one dimension of a structured representation. Language-specific grammatical realizations can then be linked to this layer as alternative or complementary ways of expressing shared functions.

To support curriculum design and language-specific implementation, Reference Level Descriptions (RLDs)[1] have been developed to specify the linguistic realizations associated with CEFR progression in individual languages. From the perspective of domain modeling, they provide a bridge from abstract proficiency descriptors to language-specific realizations that are suitable for explicit, structured encoding. For English, this bridge is made linguistically explicit in the English Profile program (Harrison and Barker, 2015), including the English Grammar Profile (EGP; O'Keeffe and Mark, 2017) and the English Vocabulary Profile (EVP; Capel, 2012), which organize grammatical and lexical features in relation to CEFR levels.

The EGP, in particular, provides a corpus-informed account of how grammatical features emerge in learner production across CEFR levels. Derived from a four-year quasi-longitudinal analysis of the Cambridge Learner Corpus, it identifies criterial features: linguistic markers that statistically distinguish adjacent proficiency levels through systematic comparison of learner and native-speaker usage. This makes it a valuable resource for grounding domain models in empirically observed relationships between linguistic realizations and proficiency development.

2.1. Graph-based domain modeling

As discussed, functional and usage-based traditions often conceptualize grammar as an interconnected system in which linguistic patterns are related through structural and functional links (Diessel, 2019). Crucially, form-function mappings are rarely one-to-one: a given form may serve multiple functions, and a given function may be realized by multiple forms. This relational organization makes grammar difficult to capture as a flat inventory of isolated items and instead motivates representations that can explicitly encode multiple types of connections.

Knowledge graphs provide such formalism by representing relational knowledge as networks of entities linked through typed relations, thereby supporting explicit organization, traversal and visualization (Peng et al., 2023). In the present case, this means representing grammatical forms, functions, and the relations between them in a machine-readable structure. As illustrated in Figure 1, such a graph can encode not only which forms realize which functions, but also how forms relate to one another through dependency and progression relations. This makes it possible to model the language domain as a structured knowledge space.

More broadly, robust domain models are essential whenever complex knowledge must be repre-

[1] https://www.coe.int/en/web/common-european-framework-reference-languages/reference-level-descriptions

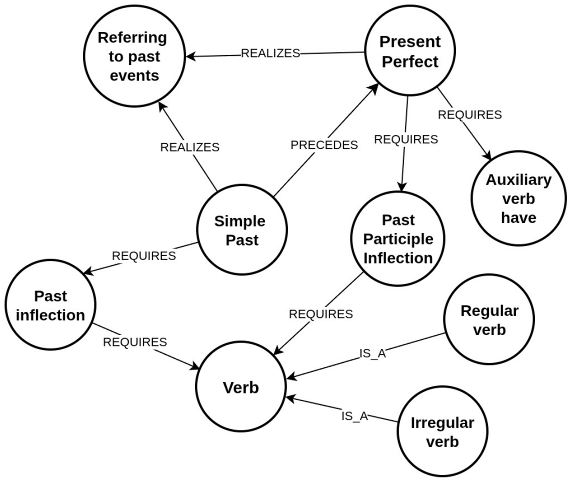

Figure 1: Illustrative example of graph-based grammar modeling. Form and Function nodes are connected through typed relations.

sented in terms of interdependent concepts and realizations (Bagchi, 2021). In exact domains such as mathematics, systems like ALEKS (Cosyn et al., 2021) demonstrate the value of relational domain modeling for adaptive learning. Grounded in Knowledge Space Theory, ALEKS organizes content through explicit prerequisite relations over large sets of topics. By contrast, ICALL has paid comparatively less attention to modeling the internal relational structure of the grammatical domain (Slavuj et al., 2017).

Recent work reflects growing interest in graph-structured approaches, though with different goals. Zhang and Li (2025), for instance, develop a dynamic vocational-English knowledge graph for real-time exercise recommendation. Vu et al. (2025) construct knowledge graphs of Russian grammar prerequisites inferred from learner performance within an ITS and evaluate their alignment with topic difficulty. In parallel, NLP-oriented work has focused more directly on formalizing linguistic knowledge itself, for example by integrating syntax with lexical-semantic networks (Prost, 2022; Rambelli et al., 2019) or by encoding grammatical constraints in ontologies (Aranovich, 2023).

Taken together, these strands point to a gap: existing graph-based approaches are typically not grounded in CEFR-based form-function mappings, do not treat pedagogical reference materials as structured linguistic data to be formalized, and are rarely released as openly reusable domain models. The present work addresses this gap by constructing a pedagogically curated, machine-readable graph of English grammar grounded in established CEFR-aligned resources.

3. The L2 Network

3.1. Implementation Considerations

The L2 Network is a typed knowledge graph that represents grammatical constructions associated with the CEFR Basic User macro-level (A1 and A2). We focus on these entry levels as a principled first step, allowing the resource to be developed and validated on a well-defined portion of the grammar domain before extension to higher proficiency bands.

The graph is implemented in Apache AGE[2], an open-source graph database built on PostgreSQL. This implementation supports the explicit encoding of typed nodes, typed edges, and metadata annotations, while also enabling graph-native querying through Cypher. As a result, the resource can be queried not only as stored data, but also as a traversable linguistic structure, for example to retrieve all forms linked to a given function, identify related or competing constructions, or trace dependency paths across relation types.

The L2 Network is publicly available on GitHub[3] under the CC BY 4.0 license. The repository includes the graph schema, machine-readable exports of nodes and edges with their properties, documentation of node and relation types, and example queries for exploration and reuse. For transparency and reproducibility, it also provides the underlying annotation-derived endorsement data together with the consensus-filtered core resource described in this paper.

3.2. Formalism and Schema

This section presents the explicit schema by which the L2 Network represents grammatical knowledge as structured data. It specifies the node types, relation inventory, and assignment criteria used to convert information from pedagogical reference materials into a typed, machine-readable graph. Formally, the resource is modeled as a directed, labeled graph $G = (V, E)$, where nodes V correspond to linguistic entities and edges E encode typed, directional relations between them. The schema is organized around two core layers, a formal layer and a functional layer, which are connected through explicitly defined relation types. Each node is assigned a unique identifier and CEFR-level metadata, and each edge contributes relational semantics to the overall structure of the domain.

[2]https://age.apache.org/

[3]https://github.com/luisards/l2-network

3.2.1. The Formal Layer (The "What")

The formal layer encodes the grammatical patterns of the domain as an explicit schema of `Form` nodes connected by typed, directed edges. It therefore constitutes the part of the resource that represents language-specific grammatical realizations in machine-readable form. Its empirical basis is the EGP, from which we adopt CEFR level assignments and ordering information. In the EGP, forms are grouped into broader grammatical supercategories, and the A1-A2 subset currently represented in the graph spans a diverse range of them, including adjectives, present clauses, modality, determiners, prepositions, noun phrases, negation, future, passives, adverbs, nouns, conjunctions, pronouns, questions, verbs, and reported speech. These source categories are retained as metadata and support both inspection of the resource and the category-level analyses reported later.

A `Form` node corresponds to an EGP element encoded under the node type *Form*. Depending on the source entry, this may represent a clause-level pattern (e.g., *SUBJ have V-en*), a morphosyntactic pattern (e.g., comparative adjective with *-er*), or a fixed lexical-grammatical unit (e.g., *What a pity!*). Each form is represented as a distinct graph node with an identifier and associated metadata, allowing heterogeneous grammatical units to be captured within a common representational framework.

Each `Form` node contains the following properties:

- `id`: internal identifier
- `label`: human-readable name
- `formula`: abstract structural pattern
- `cefr_level`: CEFR level assignment (A1 or A2)
- `egp_id`: reference to the originating EGP entry
- `example`: example sentence illustrating the form

Forms participate in multiple typed relations, modeled as many-to-many directed edges that define the topology of the formal layer. These relations are intended to make explicit different kinds of linguistic relatedness that are often only implicitly distributed across pedagogical reference materials. In the current version of the graph, three relation types are used which capture distinctions that are especially relevant for structuring the grammar domain: structural dependency, alternative realization, and progression.

Relations & assignment criteria. To make edge assignment explicit and reproducible, each relation type was defined by a separate operational criterion. `PRECEDES` edges were added only when supported by EGP ordering information. `REQUIRES` and `HAS_VARIANT` edges were added only when supported by structural evidence from reference grammars, primarily the *Cambridge Grammar of English* (Carter et al., 2016) and the British Council grammar resource (British Council).

The core relation types are defined as follows:

- `REQUIRES`: encodes structural or compositional dependency between forms. A relation `REQUIRES(A,B)` is assigned when form B cannot be adequately characterized without reference to form A as a necessary structural component. In other words, A contributes an essential part to B, and removing A would change the construction type rather than merely one of its realizations.
- `HAS_VARIANT`: links alternative realizations of a broader grammatical pattern. A relation `HAS_VARIANT(A,B)` is assigned when A and B belong to the same constructional family or instantiate the same broader grammatical pattern, but differ in morphological or structural realization, and neither is a structural prerequisite of the other.
- `PRECEDES`: encodes progression relations between closely related forms. A relation `PRECEDES(A,B)` is assigned when A and B are related constructions and the EGP places A earlier than B in its level-internal ordering, optionally supported by pedagogical grammars that present A as typically introduced before B.

These relations are treated as soft, revisable modeling assumptions rather than absolute claims about acquisition. Their role is to provide an explicit and computationally usable encoding of the internal structure of the English grammar domain, supporting inspection, traversal, and downstream pedagogical applications.

3.2.2. The Function Layer (The "Why")

The function layer encodes the communicative and semantic purposes associated with the form inventory. It provides an abstract layer of representation over language-specific grammatical realizations by modeling the kinds of meanings and discourse purposes that forms may serve. In this sense, it functions as the graph's functional ontology, aligned with the CEFR's functional-notional orientation (Green, 2012).

The function inventory was constructed by surveying CEFR A1 and A2 descriptors (Council of Europe) together with the notions and language functions associated with the relevant constructions in *Threshold Level 1990*. Overlapping or near-equivalent categories were then consolidated into a unified taxonomy, yielding a controlled inventory of function labels for use in the graph (c.f. 9 for the complete list).

Because the forms represented in the graph differ in abstraction, schematicity and lexical specificity, the associated functions also vary in granularity. In the present work, "function" is used as an umbrella term for several types of language-use categories, including communicative functions (e.g., *describing*, *reporting*), grammatical-semantic functions (e.g., *expressing degree*, *referring to entities*), pragmatic functions (e.g., *expressing politeness*, *signaling certainty or uncertainty*), and discourse-organizational functions (e.g., *introducing a topic*, *adding emphasis or contrast*). Each function was assigned a short operational definition to support consistent and inspectable mapping between forms and functions.

Each `Function` node contains the following properties:

- `function_id`: stable internal identifier
- `label`: human-readable function name
- `definition`: short operational description

The interface between the formal and functional layers is encoded through the relation `REALIZES`. A relation `REALIZES(A,B)` links a `Form` node A to a `Function` node B when the form is conventionally associated with that function in the pedagogical domain represented by the graph. The relation is not intended to claim that a form has a single inherent meaning or that the mapping is exhaustive. Rather, it encodes a pedagogically salient and linguistically motivated association between a grammatical realization and a function relevant to A1-A2 instruction and learner modeling. Because forms are often multifunctional and functions may be realized by multiple forms, `REALIZES` is modeled as a many-to-many relation. Where relevant, mappings may be further constrained to a particular CEFR level.

3.3. Structural Characteristics

To characterize the internal topology of the L2 Network, we report descriptive graph statistics capturing node distribution, relational density, and connectivity patterns in Table 1. These metrics provide a quantitative overview of the resource's structural composition and indicate that the graph forms a connected, queryable representation suitable for downstream analysis and computational use.

4. Content Validation

Our validation focused on content coverage and annotation validity, specifically on whether the form-function mappings encoded in the L2 Network adequately capture the functional scope of CEFR A1-A2 grammar. Because the form inventory itself is derived from the English Grammar Profile, the present validation targets the newly introduced functional layer rather than form inclusion. More precisely, the aim was to assess whether the communicative and semantic functions linked to each `Form` node constitute plausible and sufficiently shared interpretations of the grammatical pattern in typical A1-A2 use.

Assigning functions to grammatical forms is inherently qualitative and interpretive. Functions reflect context-sensitive communicative purposes, may overlap conceptually, vary in granularity, and admit multiple plausible analyses (Landert et al., 2023; Weisser, 2014). For this reason, expert review is standard practice in the validation of linguistically interpreted resources. In such settings, agreement metrics are best understood not as a means of eliminating all variation, but as a way of identifying the stable core of shared expert judgment. The validation procedure was therefore designed to preserve interpretive nuance while still supporting systematic resource construction: annotators could assign multiple labels and add free-text suggestions, allowing the study to capture both consensus and principled variation in the proposed mappings.

4.1. Participants and Materials

To evaluate the quality of the form-function mappings, we conducted an expert validation study with 18 annotators with backgrounds in linguistics and language teaching who were either native speakers of English or highly proficient users of English. Each annotator evaluated a set of 40 candidate items, distributed such that every item was reviewed by 2-4 annotators (mean: 2.83). For each item, annotators were asked to indicate which function(s) a given form could realize in typical A1-A2 communication.

Each annotation item included:

- the form label and formula,
- the CEFR level assignment,
- 2-3 illustrative examples, and
- a list of 5 candidate functions.

The candidate set included the target mapping, determined as described in Section 3.2.2, together with function labels drawn from the same or closely related grammatical super-categories. This design allowed us to test not only whether the proposed

Metric	Definition	Value
Total forms	Number of `Form` nodes	314
A1 forms	Number of `Form` nodes with `cefr_level = 'A1'`	119
A2 forms	Number of `Form` nodes with `cefr_level = 'A2'`	195
Functions	Number of `Function` nodes	59
Form Relations	Number of form-to-form edges	1,327
Function Relations	Number of `REALIZES` edges	600
Average degree	Mean number of functions per form (mappings / forms)	1.91
% forms with >1 function	`Form` nodes connected to more than 1 `Function` nodes	69.3%
Orphan check	Forms with neither functions nor grammatical relations	0

Table 1: Structural properties of the L2 Network.

mapping was selected, but also whether nearby alternatives were consistently preferred, thereby providing evidence about the adequacy and distinctiveness of the functional annotation layer.

4.2. Inter-Annotator Agreement

Inter-annotator agreement was assessed using three complementary metrics in order to characterize the stability of the proposed form-function annotation layer:

- **Krippendorff's α.** Because the task is multi-label, agreement was computed at the level of individual label decisions. For each standard function label, presence or absence was treated as a binary nominal variable. Krippendorff's α was used to account for chance agreement and the variable number of annotators per item (2-4).

- **Mean Jaccard similarity.** For each pair of annotators evaluating the same form, Jaccard similarity was computed as $|A \cap B|/|A \cup B|$, measuring overlap between the selected function sets.

- **Exact match rate.** This was defined as the proportion of annotator pairs who selected identical function sets for a form, representing strict set-level agreement.

Across the full annotation matrix, agreement was moderate (Krippendorff's $\alpha = 0.58$). However, aggregate agreement values mask substantial variability across forms, as illustrated in Figure 2. Upon further inspection, disagreement was primarily additive rather than contradictory: annotators typically converged on a shared core function, but differed in whether additional functions should also be assigned. This pattern is consistent with the interpretive nature of functional annotation and suggests that the main source of variation lies in the breadth of mappings retained, rather than in direct opposition over a form's primary function.

4.3. Expert Endorsement Rate

To quantify support for individual form-function mappings, we computed an endorsement rate for each pair, defined as the proportion of experts assigning a given function to a given form. Let n_s denote the number of experts selecting the function and n the total number of experts evaluating that item. The endorsement rate r is given by:

$$r = \frac{n_s}{n}$$

Endorsement rates were computed for every candidate form-function pair in order to examine the distribution of consensus across the annotation layer. This allows us to distinguish highly stable mappings from those that attract weaker or more divided support, and to better understand whether disagreement reflects direct contradiction or the addition of secondary functions.

Endorsement band	Vote patterns	Mappings
Unanimous	$2/2,\ 3/3,\ 4/4$	309
75% annotators	$3/4$	18
67% annotators	$2/3$	76
50% annotators	$2/4,\ 1/2$	133
Below 50%	$1/3,\ 1/4$	242

Table 2: Distribution of endorsement levels across candidate form-function pairs by attainable vote pattern.

As shown in Table 2, most candidate pairs received either unanimous endorsement or majority support, while a substantial minority fell below the 50% level. In line with the agreement analysis above, these lower-consensus cases often reflect additive disagreement: annotators typically converged on a shared core function, but differed in whether additional functions should also be assigned. Endorsement rates therefore provide a useful descriptive view of which parts of the functional layer are especially stable and which remain more weakly supported.

145 forms achieved both monofunctionality and unanimous endorsement, representing the highest-confidence mappings in the resource. These

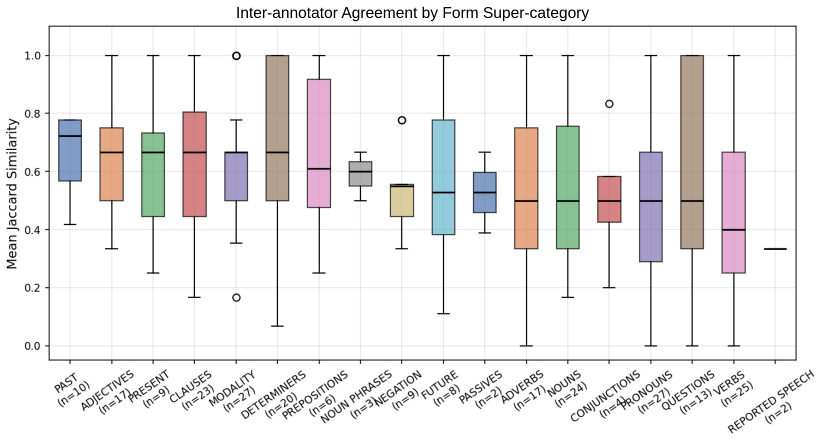

Figure 2: Distribution of inter-annotator agreement across forms' super-categories.

cases can be interpreted as especially stable form-function associations within the current annotation scheme. Examples are shown in Table 3.

4.4. Consensus Analysis and Resource Release

Because the number of annotators per form varied between two and four, the same endorsement percentage does not always correspond to the same evidentiary standard. To derive a stable core resource for public release, we therefore adopted an *independent corroboration* criterion: a mapping is retained only if it was endorsed by at least two annotators ($n_{\text{sel}} \geq 2$).

This criterion applies a consistent principle across the dataset: each released mapping must be supported by agreement between at least two independent judges, even though the equivalent endorsement rate differs by annotator group size (Table 4).

Applying this filter yielded 429 mappings, reducing the original inventory by 171 mappings (28.5%) and 2 forms. The two excluded forms, *Pronoun subject it for first person* and *Verb infinitive form*, were cases in which annotators selected entirely different functions, indicating that these entries require further adjudication. We therefore release both the full endorsement data and the independently corroborated subset, preserving lower-consensus mappings for future analysis while providing a principled consensus-based core annotation layer for reuse.

4.5. Operational Integration

To demonstrate runtime usability, we integrated the L2 Network into AISLA (Chen et al., 2022), a task-based dialogue system for curriculum-aligned English practice. Within this architecture, the L2 Network functions as a queryable domain model that determines which forms can be selected, how they map to functions, and under what constraints they are instantiated.

4.6. Implementing a Contextual Layer (The "How")

While the formal and functional layers represent what linguistic knowledge is modeled, instructional systems must also specify how that knowledge is practiced. In task-based language learning, AISLA's pedagogical framework, forms are practiced through interactional situations in which particular form-function pairings become meaningful and pragmatically licensed. To capture this contextual layer, we introduce a `Task` node type connected to `Form` and `Function` nodes via the typed edges `TARGETS_FORM` and `TARGETS_FUNCTION`. An example of these layers is shown in Figure 3.

Each `Task` node represents a parameterized dialogue template grounded in a common real-life interactional scenario that pragmatically licenses one or more target forms and functions. In this way, the graph does not only represent what linguistic knowledge is modeled, but also how that knowledge can be contextualized for instruction. Each `Task` node contains the following properties:

- `id`
- `task_name`
- `form_id`
- `function_id`
- `system_instructions`
- `student_instructions`

Form	Formula	Function
AffirmativeDeclarativeClause	`SUBJ V (...)`	Making statements
AffirmativeInterrogativeClauseWithBe	`(wh) BE SUBJ (...)`	Asking for information
AttributiveAdjective	`ADJ N`	Describing
FutureWillAffirmative	`SUBJ will V`	Referring to future actions and events
QuantifierWithPluralNoun	`QUANT N.PL`	Expressing quantity
DeterminerWithNoun	`DET N`	Referring to entities
TimeAdverbEnd	`CLAUSE ADV.TIME`	Situating events in time
BeSureClause	`be sure that CLAUSE`	Expressing certainty

Table 3: Examples of unanimously endorsed, single-function mappings.

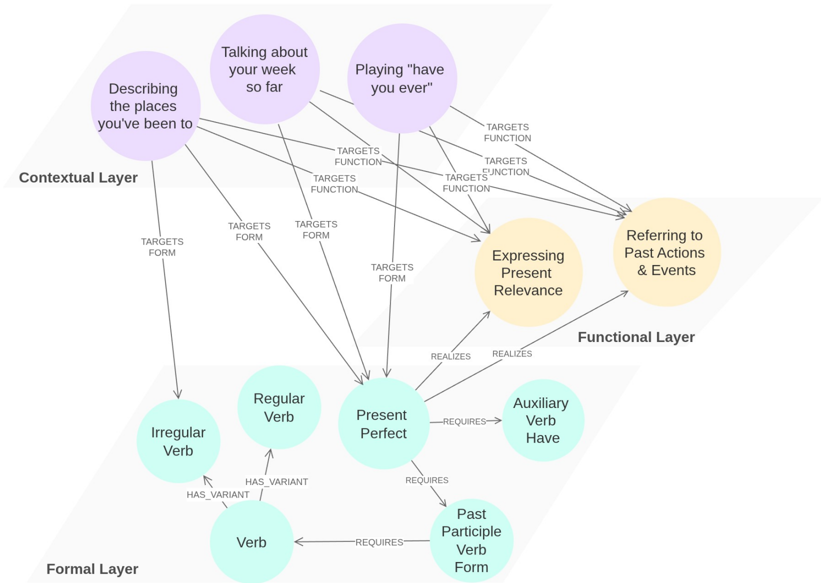

Figure 3: Example of the contextual extension layer in the L2 Network. Task nodes connect pedagogical contexts to the underlying function and form layers.

Annotators	Min. support ($n_{\text{sel}} \geq 2$)	Rate
2	$2/2$	100%
3	$2/3$	67%
4	$2/4$	50%

Table 4: Equivalent endorsement rates for the independent corroboration criterion across annotator group sizes.

Because task templates are linked to forms and functions through graph relations rather than hard-coded directly into prompts, linguistic targeting and contextual realization can be handled separately. This makes it possible to vary instructional contexts while preserving explicit control over what is being practiced.

The graph is particularly relevant for learner modeling because it provides an explicit structure over which a learner overlay can be defined. Rather than tracking performance only on isolated items, the system can associate learner evidence with `Form` and `Function` nodes and interpret this evidence in relation to neighboring nodes. This makes it possible to represent learner state not simply as a list of correct or incorrect targets, but as a structured profile over the grammar domain.

Such an overlay can support more principled adaptive decisions. Depending on the learner's observed performance, the system may identify forms or functions that appear underdeveloped, revisit prerequisite structures, remain within the same functional area while shifting to a related form, or select practice that reinforces a weak form-function mapping. In this way, the graph supports decisions about what content should be delivered next. Rather than relying on an LLM to infer pedagogical priorities, the the addition of the graph makes the relevant domain structure explicit and queryable.

5. Conclusion & Future Work

This paper introduced the L2 Network, a graph-based English grammar domain model that links CEFR-aligned grammatical forms to functions via explicit relations. By making linguistic targets and their dependencies machine-readable, this resource provides a transparent substrate for principled sequencing, learner-model overlays, and integration with LLM-based generation. A central advantage of the proposed schema is its extensibility. It distinguishes a stable grammatical core from additional system layers, enabling platform-specific adaptation without modifying the underlying form-function mappings.

The analysis indicates meaningful expert convergence on the functions of Basic User-level forms, while also reflecting the inherently interpretive nature of form-function mapping. The observed Krippendorff's α of 0.58 suggests a shared but still evolving annotation space rather than a fully stabilized inventory. At the same time, the substantial proportion of unanimously endorsed mappings indicates that the current version of the graph already captures a stable core of form-function associations. We therefore treat these results as encouraging initial validity evidence, while recognizing that the resource remains a work in progress and will require continued refinement and evaluation.

Several directions follow: First, we plan to extend the graph with lexical realizations, supporting more reliable controlled text generation by conditioning LLM outputs on explicit specifications. Second, future work will extend the present content-validation study toward predictive and external criterion validation. In particular, we will examine whether graph-defined form-function mappings and level assignments predict learner production patterns and CEFR-aligned proficiency outcomes in independent datasets.

Finally, although instantiated here for English, the design is intended to be transferable across languages. The CEFR provides a shared, language-agnostic functional backbone, while RLDs supply language-specific realizations at each level. For languages with comparable RLD resources, the L2 Network offers a generalizable blueprint for constructing structured, data-informed domain models that can support transparent adaptivity in ICALL.

Beyond domain modeling, the L2 Network provides a symbolic substrate for neuro-symbolic approaches to GenAI, curriculum and syllabus design as well as assessment design and test coverage analysis. Because constructions (forms-function pairings) and their dependencies are represented explicitly, the graph can support symbolic planning of pedagogical intent (e.g., selecting function-consistent targets under soft prerequisite and competition constraints) followed by neural realization via LLM generation. It also enables constraint- and verification-based workflows in which generated dialogues are automatically checked against graph-derived expectations and iteratively refined. Furthermore, graph neighborhoods can be retrieved as structured context to condition generation (structure-aware retrieval), or translated into soft or hard decoding constraints to improve controllability.

6. Acknowledgements

We thank the researchers who participated in the expert survey and contributed their time and expertise to the validation of the function mappings. Their feedback was essential for assessing the clarity and consistency of the proposed annotations.

We also acknowledge the use of the *English Grammar Profile* as the primary resource for the form inventory and CEFR-level anchoring used in this work.

This work was supported by the Hector Foundation (Pillar III).

7. Limitations

A primary limitation of the current resource lies in the compression of annotator judgments into a filtered consensus graph. While this step improves consistency and yields a more stable set of mappings for downstream implementation, it necessarily suppresses minority selections that may reflect context-sensitive or more fine-grained functional interpretations. The filtered inventory is therefore likely more robust for modeling and pedagogical control, but less exhaustive as a representation of the full functional variability associated with the constructions.

A second limitation concerns scope. The current version of the L2 Network covers only English grammar at the CEFR Basic User levels (A1-A2). Although this restriction was intentional in order to support careful schema development and validation, it means that the present resource does not yet capture higher proficiency levels, broader discourse phenomena, or cross-linguistic variation.

A further limitation is that several aspects of the graph depend on expert modeling decisions. This applies not only to the assignment of form-function mappings, but also to the granularity of the function inventory and to the operational criteria used to define relations. These choices were made transparently and grounded in established reference materials, but they remain revisable as the resource evolves.

Finally, the validation reported here is limited to content coverage, agreement patterns, and structural consistency. Although these provide important initial evidence for the usefulness of the resource, they do not by themselves establish the effectiveness of the graph in downstream applications such as learner modeling, adaptive task selection, or controlled generation. Such uses require separate empirical evaluation.

8. Bibliographical References

Raúl Aranovich. 2023. Modeling grammars with knowledge representation methods: Subcategorization as a test case. In *European Semantic Web Conference*, pages 112–116. Springer.

Mayukh Bagchi. 2021. A large scale, knowledge intensive domain development methodology. *Knowledge Organization: KO*, 48(1):8.

Heike Behrens. 2009. Usage-based and emergentist approaches to language acquisition.

Luca Benedetto, Gabrielle Gaudeau, Andrew Caines, and Paula Buttery. 2025. Assessing how accurately large language models encode and apply the common european framework of reference for languages. *Computers and Education: Artificial Intelligence*, 8:100353.

Conrad Borchers and Tianze Shou. 2025. Can large language models match tutoring system adaptivity? a benchmarking study. In *International Conference on Artificial Intelligence in Education*, pages 407–420. Springer.

British Council. Grammar. LearnEnglish. Accessed: 2026-02-17.

Annette Capel. 2012. Completing the english vocabulary profile: C1 and c2 vocabulary. *English Profile Journal*, 3:e1.

Ronald Carter, Michael McCarthy, Geraldine Mark, and Anne O'Keeffe. 2016. English grammar today: An a–z of spoken and written grammar. Web adaptation in Cambridge Dictionary. Accessed: 2026-02-17.

Xiaobin Chen, Elizabeth Bear, Bronson Hui, Haemanth Santhi-Ponnusamy, and Detmar Meurers. 2022. Education theories and ai affordances: Design and implementation of an intelligent computer assisted language learning system. In *International Conference on Artificial Intelligence in Education*, pages 582–585. Springer.

Eric Cosyn, Hasan Uzun, Christopher Doble, and Jeffrey Matayoshi. 2021. A practical perspective on knowledge space theory: Aleks and its data. *Journal of Mathematical Psychology*, 101:102512.

Council of Europe. 2001. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion Volume*. Cambridge University Press, Cambridge.

Holger Diessel. 2019. *The Grammar Network*. Cambridge University Press.

Holger Diessel. 2023. *The constructicon: Taxonomies and networks*. Cambridge University Press.

Adele E Goldberg. 2003. Constructions: A new theoretical approach to language. *Trends in cognitive sciences*, 7(5):219–224.

Anthony Green. 2012. *Language functions revisited: Theoretical and empirical bases for language construct definition across the ability range*, volume 2. Cambridge University Press.

Julia Harrison and Fiona Barker. 2015. *English profile in practice*, volume 5. Cambridge University Press.

Kristiina Jokinen. 2024. The need for grounding in llm-based dialogue systems. In *Proceedings of the Workshop: Bridging Neurons and Symbols for Natural Language Processing and Knowledge Graphs Reasoning (NeusymBridge)@ LREC-COLING-2024*, pages 45–52.

Anisia Katinskaia. 2025. An overview of artificial intelligence in computer-assisted language learning. *arXiv preprint arXiv:2505.02032*.

Daniela Landert, Daria Dayter, Thomas C. Messerli, and Miriam A. Locher. 2023. *Corpus Pragmatics*. Cambridge Elements in Pragmatics. Cambridge University Press.

Diane Larsen-Freeman. 2003. Teaching language: From grammar to grammaring. *(Heinle Cengage Learning)*.

Anne O'Keeffe and Geraldine Mark. 2017. The english grammar profile of learner competence: Methodology and key findings. *International Journal of Corpus Linguistics*, 22(4):457–489.

Indira Padayachee et al. 2002. Intelligent tutoring systems: Architecture and characteristics. In *Proceedings of the 32nd Annual SACLA Conference*, pages 1–8. South African Computer Lecturers' Association.

Ciyuan Peng, Feng Xia, Mehdi Naseriparsa, and Francesco Osborne. 2023. Knowledge graphs: Opportunities and challenges. *Artificial intelligence review*, 56(11):13071–13102.

Jean-Philippe Prost. 2022. Integrating a phrase structure corpus grammar and a lexical-semantic network: the holinet knowledge graph. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 613–622.

Giulia Rambelli, Emmanuele Chersoni, Philippe Blache, Chu-Ren Huang, and Alessandro Lenci. 2019. Distributional semantics meets construction grammar. towards a unified usage-based model of grammar and meaning. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 110–120.

Hans-Jörg Schmid. 2015. A blueprint of the entrenchment-and-conventionalization model. *Yearbook of the German cognitive linguistics association*, 3(1).

Vanja Slavuj, Ana Meštrović, and Božidar Kovačić. 2017. Adaptivity in educational systems for language learning: a review. *Computer Assisted Language Learning*, 30(1-2):64–90.

Jan Ate Van Ek, John Leslie Melville Trim, et al. 1991. *Threshold level 1990*. Council of Europe.

Anh-Duc Vu, Jue Hou, Anisia Katinskaia, Ching-Fan Sheu, and Roman Yangarber. 2025. A bayesian approach to inferring prerequisite structures and topic difficulty in language learning. In *BEA: Proceedings of the 20th Workshop on Innovative Use of NLP for Building Educational Applications*.

Martin Weisser. 2014. *Speech act annotation*, page 84–114. Cambridge University Press.

Fengna Zhang and Xinxin Li. 2025. Dynamic recommendation system for higher vocational english learning paths based on real-time knowledge graph update. *Engineering Reports*, 7(12):e70492.

9. Appendix

ID	Function	ID	Function
1	Asking for advice	31	Giving advice
2	Asking for confirmation	32	Giving directions or instructions
3	Asking for information	33	Making offers and invitations
4	Comparing	34	Making suggestions
5	Comparing within a set	35	Providing detail
6	Conveying meaning	36	Referring back to entities
7	Making statements	37	Referring to actions states or relations
8	Describing	38	Referring to additional or alternative entities
9	Emphasizing the affected entity	39	Referring to entities
10	Expressing obligation or necessity	40	Referring to future actions and events
11	Expressing beliefs and opinions	41	Referring to impersonal events
12	Expressing cause	42	Referring to past actions and events
13	Expressing certainty	43	Referring to past events with present relevance
14	Expressing condition	44	Referring to present states events and facts
15	Expressing degree	45	Referring to self
16	Expressing evaluation	46	Referring to self-directed actions
17	Expressing frequency	47	Referring to the acting or experiencing entity
18	Expressing hypotheses or inferences	48	Referring to the affected entity
19	Expressing manner	49	Referring to third-person subjects
20	Expressing negation	50	Reporting information
21	Expressing politeness	51	Requesting
22	Expressing possession	52	Sequencing events
23	Expressing possibility	53	Situating events or entities
24	Expressing ability	54	Situating entities in space
25	Expressing preference	55	Situating events in time
26	Expressing probability	56	Specifying information
27	Expressing progression	57	Pointing to entities
28	Expressing quantity	58	Expressing contrast
29	Expressing wishes	59	Expressing indefiniteness
30	Framing or linking information		

Table 5: Inventory of communicative functions ($n = 59$).