MultiCoS: A Multilingual Dataset of Connective Semantics with Context–Sentence Compatibility

Anne Mucha, Ciyang Qing, Wataru Uegaki
School of Philosophy, Psychology and Language Sciences, University of Edinburgh
3 Charles Street, Edinburgh, EH8 9AD United Kingdom
{amucha, cqing, w.uegaki}@ed.ac.uk

Abstract

We present a multilingual dataset of connective semantics. The dataset contains the semantic annotations of clausal connectives (e.g. *and* and *or* in English) from 24 languages, based on our original native-speaker elicitation data. Unlike existing lexica on connectives, the dataset includes systematic evidence for the annotations in the form of context-sentence compatibility judgments, including negative evidence. The paper describes the methodology of data collection and the format of the dataset. We also discuss its potential use cases for the validation of cross-linguistic generalizations, examinations of their potential counterexamples, and for benchmarking felicity judgments by NLU systems.

Keywords: connectives, coordination, semantics, multilingual dataset, lexicons, typology, cross-linguistic generalizations, cross-linguistic variations

1. Introduction

In this paper, we describe our multilingual dataset of connective semantics.[1] The semantics of connective coordinators, such as *and* and *or* in English, exhibit a number of cross-linguistic generalizations and rich cross-linguistic variations (Haspelmath, 2007). At the same time, modern semantic theories have enabled fine-grained analyses of connective meanings cross-linguistically (e.g. Winter, 2002; Bowler, 2014; Mitrovic and Sauerland, 2014; Flor et al., 2017). For these reasons, connectives offer an ideal testing ground for theories of language universals/variation within the functional/logical vocabulary (e.g., Horn, 1972; Katzir and Singh, 2013; Uegaki, 2023; Bar-Lev and Katzir, 2025).

For NLP, inference arising from coordination is one of the key lexico-syntactic phenomena for the improvement of Natural Language Understanding (NLU) and Recognizing Textual Entailment (RTE) systems (Sammons, 2015). Accurate interpretation of clausal coordination remains a persistent challenge for NLU models, as they often fail to capture the compositional semantics of connectives such as *and* and *or*, especially concerning non-Boolean aspects of their meanings (Naik et al., 2018; Glockner et al., 2018; Saha et al., 2020).

By providing systematically annotated judgments of compatibility between coordinated sentences and contexts across 24 languages, our dataset offers a linguistically grounded resource for probing and benchmarking cross-linguistic understanding of coordination in linguistics/cognitive science as well as in NLU models. It thus serves both as an empirical basis for research on the semantics of connectives and as a diagnostic benchmark for multilingual entailment and reasoning tasks.

p	q	AND	OR	NOR	NAND	XOR
T	T	T	T	F	F	F
T	F	F	T	F	T	T
F	T	F	T	F	T	T
F	F	F	F	T	T	F

Table 1: The truth table for the Boolean semantics of connectives in the square of opposition + XOR

2. Background

2.1. Semantic properties of connectives

In our dataset, we annotate the following semantic properties of connectives.

First, in terms of their Boolean truth conditions, a binary connective C maps two truth values to a truth value. Out of the $2^4 = 16$ possible functions, AND,[2] OR, NOR, and NAND (whose truth conditions are shown in Table 1), are called *the square of opposition*. Note that p *OR* q is defined to be true when p and q are both true, even though it may pragmatically imply that p and q are not both true. This *inclusive disjunction* OR contrasts with the *exclusive disjunction* XOR, which is simply false when its two arguments are both true.

[1] The dataset can be accessed at https://github.com/EdinburghMeaningSciences/multicos.

[2] We use CAPS (as in AND and OR) to denote connectives in the semantic metalanguage, which may be used to describe the semantics of natural language connective expressions, which are denoted by *italics* (as in *and* and *or*).

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 4861–4871
11–16 May 2026.

Our annotation also considers non-Boolean aspects of connective semantics. In particular, some connectives may mark a contrast between their two arguments. For instance, p *BUT* q not only asserts that both p and q are true, but also implicates a contrast between the two.

Furthermore, disjunctive connectives in questions $?(pCq)$ may give rise to two types of readings. The *polar question* (PolQ) reading asks whether the result of applying the connective is true (and therefore expects a Yes/No answer), whereas the *alternative question* (AltQ) reading asks which disjunct is true (and therefore expects the answerer to identify the true disjunct). In languages such as English, the same disjunctive connective *or* is used for both readings (1), with different intonational patterns employed to disambiguate them.

(1) Do you want tea or coffee? (✓AltQ, ✓PolQ)

Other languages such as Basque have disjunctive connectives that exclusively give rise to an AltQ reading (2) or a PolQ reading (3) (Saltarelli, 1988).

(2) Te-a ala kafe-a nahi duzu?
tea-Art or coffee-Art want you.it
"Which one is true: you want tea or you want coffee?" (✓AltQ, *PolQ)

(3) Te-a edo kafe-a nahi duzu?
tea-Art or coffee-Art want you.it
"Is it true that you want tea or coffee?"
(*AltQ, ✓PolQ)

Finally, disjunction under a possibility modal may give rise to a *free-choice* inference exemplified in (4), an inference that is unavailable under the Boolean semantics for OR, (5) (Kamp, 1973).

(4) You may have tea **or** coffee.
$\rightsquigarrow$ You may have tea **and** you may have coffee.

(5) p OR $q \not\rightsquigarrow p$ AND q

2.2. Existing generalizations and typology

Out of the many logically possible connectives, only a subset are lexicalized in natural language. Special attention in the literature has been paid to lexicalization patterns with regard to the meanings represented in Table 1. The following generalizations are particularly relevant for our purposes (see Horn 1972, Katzir and Singh 2013): while languages often lexicalize AND and OR, NOR is less likely to be lexicalized as a morphologically simple connective, and NAND is not lexicalized at all in natural languages. Notably, however, not all languages have dedicated connectives for conjunction (AND) or disjunction (OR). Instead, languages may use simple juxtaposition to express conjunction, as well as modal, conditional or interrogative constructions to express disjunction, without any overt connectives (Haspelmath, 2007; Mauri, 2008). It has also been proposed that languages may use a connective that semantically encodes disjunction (OR) to express conjunctive meaning, when it undergoes obligatory pragmatic strengthening due to the absence of a connective encoding AND (Bowler, 2014).

Some interesting typological generalizations have been proposed for natural language disjunction in particular. For instance, a typological survey by Mauri (2008) illustrates that many languages use some sort of irrealis marking (e.g. markers of modality, non-factive mood or conditional structures) to express disjunction. This is because disjunction essentially conveys the possibility of different alternatives being true (see also Zimmermann 2000 for a semantic analysis of disjunction along these lines). Based on a sample of 60 languages, Mauri (2008) generalizes that all constructions that express disjunctive meaning in natural language contain either an overt disjunctive connective or irrealis marking, or a combination of both. A consequence of this generalization is that simple juxtaposition of coordinated sentences cannot be used to express disjunctive meaning.

Another relevant generalization is that natural languages do not lexicalize the distinction between inclusive and exclusive disjunction, as stated, e.g. in Haspelmath (2007) and Katzir and Singh (2013). According to the literature, disjunctive connectives always realize OR, which semantically encodes inclusive disjunction but is often pragmatically strengthened to receive an exclusive interpretation. However, there are no connectives that lexically encode exclusive disjunction (XOR).[3]

2.3. Existing language resources

There are several existing language resources that describe connective semantics. They can be divided into two types: lexicons and grammar databases. The relevant lexicons include Connective-Lex (Stede et al., 2019), Turkish Connectives Lexicon (Zeyrek and Başıbüyük, 2019), DimLex-Eng (Das et al., 2018), Naija-Lex (Marchal et al., 2021), Chinese-DiMLex (Wan et al., 2023) and Thai Discourse Treebank (Prasertsom et al., 2024). These resources describe the inventories of connectives in the sample languages with information of their syntactic categories and

[3]Some complex disjunctions, such as German *entweder ... oder* or French *soit ... soit* tend to be perceived as "more exclusive" than their simple counterparts. It has been argued that the pragmatic inference that leads to an exclusive interpretation is close to obligatory with these elements. Still, they do not semantically encode exclusive disjunction (see e.g. Spector 2014).

semantic/pragmatic sense annotations. Grammar databases include WALS (Haspelmath et al., 2005), Grambank (Skirgård et al., 2023) and Terraling (Koopman and Guardiano, 2022), which describe per-language features of grammatical properties relevant for coordination.[4] The information on the sample languages and the types of these resources are summarized in Table 2.

Our dataset is different from these existing resources in important respects. We first highlight two differences from the lexicon-type resources surveyed above: (a) Compared to the lexicons surveyed, the language coverage of our dataset is broader, containing data from a typologically diverse set of 24 languages assembled through a uniform data collection methodology; (b) Unlike the lexicons, our dataset systematically contains examples that support the semantic properties of connectives, in the form of compatibility between specific contexts and coordinated sentences. Crucially, these examples also include *negative evidence*, where a sentence is judged to be infelicitous given a context. In this sense, our data are comprehensive and support reproduction of conclusions regarding the senses of individual connectives.

The grammar databases have broader language coverages compared to our dataset. However, the data in these databases are different in nature from the data in our dataset in two crucial respects: (a) the databases record specific *per-language* features, and do not systematically describe properties of *individual connectives* based on linguistic examples; (b) the features recorded are largely morpho-syntactic in nature (e.g., formal similarities between conjunctions and universal quantifiers; see fn. 4) and do not systematically contain semantic information. Terraling is the closest to our dataset in that it contains linguistic examples to support feature values. However, since these examples are not paired with felicity judgments with respect to varying contexts, they are not designed to support conclusions about the range of senses the relevant connectives are compatible with. Thus, our dataset is the first multilingual dataset that offers information of connective semantics with comprehensive supporting data in the form of context-sentence compatibility.

[4] Relevant features from WALS, Grambank and Terraling include the following:

- WALS 56A: Formal similarities between conjunctions and universal quantifiers
- WALS 64A: Formal identity/differentiation between nominal and verbal Conjunction
- Grambank GB027: Are nominal conjunction and comitative expressed by different elements?
- Terraling 01: Coordination of proper names
- Terraling 101: Coordination of VP-predicates

3. Methodology of data collection

3.1. Language sample

In (6), we list the languages for which we elicited data on natural language connectives, grouped by language family where applicable.

(6) **Sample languages:** Dutch, Farsi, Greek, Hindi, Russian, Spanish (Indo-European), Vietnamese, Khmer (Austro-Asiatic), Thai (Kra-Dai), Cantonese, Mandarin (Sino-Tibetan), Hausa, Hebrew (Afro-Asiatic), Hungarian (Uralic), Akan, Igbo, Kîîtharaka (Niger-Congo), Tagalog (Austronesian), Turkish (Turkic), Telugu (Dravidian), Basque, Japanese, Korean, Mapudungun

The language sample was constructed such as to strike a balance between typological diversity and feasibility. We focused on languages that were accessible to us via existing contacts to native speakers. However, we aimed to cover as many major language families as possible as well as the main subgroups of the Indo-European family.

3.2. Questionnaire and elicitation procedure

Data elicitation was conducted online via video conference calls, with one native speaker consultant per language. The consultants were compensated for their time and effort (9 GBP/hour). For every language, we split the data collection into two sessions with separate questionnaires. The first questionnaire was designed to elicit the set of connectives that are canonically used in conjunctive and disjunctive contexts, which we manipulated in ways that are known to influence the choice of connective in some languages (see e.g. Haspelmath 2007). This was implemented with *translation in context*-tasks. Specifically, we presented speakers with contexts such as (7), and with English target sentences such as (7-a), which we asked the consultants to translate into their native language. By way of example, (7-b) shows a translation of the target sentence in Thai.

(7) *Context (disjunction)*: Your sister asks you what Paul and John do for a living. You know one of them is a doctor but you don't know which one. You say: ...
   a. *Target sentence:* Paul is a doctor or John is a doctor.
   b. *Thai translation:*
      Paul pen mŏr **rŭe** John pen mŏr.
      Paul be doctor OR John be doctor

The context in (7) was designed to elicit a disjunctive connective (i.e. the equivalent of English *or*).

Resource	Languages covered	Type
Connective-Lex (Stede et al., 2019)	Arabic, Bangla, Czech, Dutch, English, French, German, Italian, Portuguese	Lexicon
DimLex-Eng (Das et al., 2018)	English	Lexicon
Turkish Connectives Lexicon (Zeyrek and Başıbüyük, 2019)	Turkish	Lexicon
Naija-Lex (Marchal et al., 2021)	Nigerian Pidgin	Lexicon
Chinese-DiMLex (Wan et al., 2023)	Mandarin Chinese	Lexicon
Thai Discourse Treebank (Prasertsom et al., 2024)	Thai	Lexicon
WALS (Haspelmath et al., 2005)	∼300 languages	Grammar database
Grambank (Skirgård et al., 2023)	1,778 languages	Grammar database
Terraling (Koopman and Guardiano, 2022)	∼30 languages	Grammar database

Table 2: The list of existing language resources concerning multilingual connective semantics. The languages covered column for the grammar databases are shows the number of languages covered in the portion of the database that pertain to the connective semantics.

Additional example contexts in the domain of disjunction were constructed to manipulate i) whether the disjunctive interpretation is exclusive or inclusive and ii) whether or not the speaker knows which disjunct is true. Moreover, we elicited disjunctive connectives in questions, in the scope of negation, and in free choice contexts.

Conjunctive connectives were elicited in contexts with and without a contrastive meaning component (targeting the difference between BUT and AND). In addition, we elicited contrastive conjunction with one conjunct negated (e.g. *Paul is not a doctor but [he is] a receptionist*), as well as the connective strategies that languages use in contexts in which both connected clauses are false (targeting NOR).

Our second questionnaire consisted of felicity judgment tasks as standardly employed in semantic fieldwork (see e.g. Matthewson 2004, 2011). The aim of this questionnaire was to investigate the range of conjunctive and disjunctive meanings that the connectives in our sample languages can express. For this purpose, we constructed new contexts that fit the target sentences elicited in the first questionnaire, but trigger different interpretations than the context that was initially used for elicitation. For instance, we presented the Thai sentence elicited in (7) in the context in (8), which was constructed to test whether an *inclusive* reading of the disjunction is possible. The consultant's task was to judge the sentence's felicity in the given context.

(8) *Context (disjunction, inclusive)*: Your sister asks you what Paul and John do for a living. You know at least one of them is a doctor. You can't quite remember though, maybe John is a doctor, maybe Paul is a doctor, maybe both of them are doctors. You say: ...

Paul pen mŏr **rŭe** John pen mŏr.
Paul be doctor OR John be doctor

In addition, we have recorded the felicity of juxtaposition examples without any overt connective marking given the contexts used in the first questionnaire. The data collected thus also supports testing of cross-linguistic hypotheses concerning juxtaposition (Haspelmath, 2007; Mauri, 2008).

This data collection methodology has some specific advantages, namely: i) The same questionnaire was used for every language in our sample, allowing for systematic cross-linguistic comparison. ii) In addition to production data from the translation tasks, felicity judgments in context reveal what kind of meaning can *not* be expressed by a particular connective (i.e. negative evidence), thus facilitating semantic analysis. iii) Systematic combination of conjunctive/disjunctive contexts with all elicited connectives, as well as juxtaposition, results in a comprehensive overview of the connectives' possible uses, and allows to identify multifunctional connectives within the sample languages, as well as cross-linguistic patterns in form-meaning-mappings.

4. Structure and format of the dataset

The dataset can be accessed at `https://github.com/EdinburghMeaningSciences/multicos`. The format of the data for each language consists of the metadata YAML file and three CSV tables: the examples table, the evidence table, and the summary table. In a nutshell, **the examples table** lists all relevant example sentences while **the evidence table** records felicity judgments regarding the examples in the example table relative to various

contexts. **The summary table** summarizes the inventory of connectives, with information about morpho-syntactic and semantic properties of each connective in the language. In addition, the repository contains the language-general **questionnaire table** (`connectives_questionnaire.csv`), which contains the contexts used to obtain the felicity of examples sentences. Below, we describe the four language-specific data files.

Language metadata (`lang-metadata.yml`)
The metadata YAML file contains the following information about the language: language name, ISO 639-3 code, Glottolog code, consultant demographics, and elicitation dates.

Example table (`lang-examples.csv`)
The example table contains all examples elicited for the language with unique reference for each example. The examples are represented in a way similar to the Interlinear Glossed Text. For each row, the value in the **words** column corresponds to the transcription with word boundaries; the the value in the **glosses** column corresponds to the word-by-word glosses; and the values in the **translation** column corresponds to the English translation of the example. See Table 4 for a subpart of the example table for Kîîtharaka.

For each example, two forms of the relevant connective are identified: the **expression** core form and its **full form**. The full form of a connective consists of morphologically independent elements that linearly intervene between the coordinated clauses and/or particles that adjoins to one or more of the clauses. If the full form consists of one word, the full form and the expression core form are identical. If the full form consists of multiple morphosyntactic elements, we identified the core element among them, according to the following constraints: (a) choose an element that is more specific in semantic coverage (in terms of the contexts it is compatible with) as the core expression; (b) when choosing between a particle that attaches to coordinated clauses and an element that intervenes between clauses, choose the latter as the core element.

Evidence table (`lang-evidence.csv`)
The evidence table records the acceptability and felicity judgments of the examples given a certain context in the questionnaire table. For example, in the part of the evidence table for Kîîtharaka in Table 4, the initial row indicates that example `kii_1` in the example table is *felicitous* in context `conj-nocontrast-sta` (a noncontrastive conjunctive context with statives), which is in the questionnaire table. This piece of evidence then has the unique reference `kii_ev_1`. On the other hand, the same example is *infelicitous* in context `disj-nspk-sta-exc-2` (a disjunctive context without speaker knowledge with statives). As these cases illustrate, the same example can be judged to be felicitous or infelicitous depending on the context. This dataset format thus contains systematic and reproducible evidence concerning the range of contexts example sentences are compatible with.

Summary table (`lang-summary.csv`)
The summary table records the properties of each connective in the language, given the evidence available in the evidence table. Each row in the table corresponds to an observation regarding whether a specific connective or juxtaposition is compatible with a certain context type (characterized by semantic/pragmatic features), with reference to specific pieces of evidence in the evidence table. For example, in the subpart of the summary table for Kîîtharaka in Table 4, the initial row indicates that juxtaposition is compatible with ('can_express') a non-contrastive conjunctive context type with eventives, with reference to evidence `kii_ev_3` and `kii_ev_30` in the evidence table. The features below define the context types:

- **kboth**: the speaker's belief about the conjunction of clauses;
- **kneither**: the speaker believes neither clause;
- **contrast**: there is a contrast between the clauses;
- **stative**: the clauses describe states rather than dynamic events;
- **negated_p**: the first clause has overt negation;
- **Kp**: for either clause, the speaker believes its truth;
- **question**: the sentence is a question;
- **fc**: free choice inference is licensed

Aside from **kboth**, the features are valued with 1 or 0. **kboth** is valued as 1, 0, or '?'. The '?' value is used to specify the speaker's uncertainty as in the context in (8).

The summary table also includes information on whether the full form of the connective morphosyntactically contains **negation**, and if so, whether it appears syntactically *above* the coordination (as in NEG[p C q]) or *below* (as in [NEGp C NEGq]).

5. Case studies

5.1. Testing generalizations

Based on the dataset, we tested the four cross-linguistic generalizations discussed in Sect. 2.2:

1. There is no lexicalization for NAND;
2. NOR is unlikely to be lexicalized as a morphologically simple connective;

Ref	Expression	Full_Form	Words	Glosses	Translation
kii_1	na	na	Susie ni n-dagitaarî na ni mû-arimû	Susie foc 9-doctor and foc 1-teacher	Susie is a doctor and she is a teacher.
kii_2	na	na	Paul ni-a-kû-on-a TV na ni-a-kû-baac-a n-guo	Paul foc-1sm-pres-see-fv TV and foc-1sg.sm-pres-iron-fv 10-clothes	Paul is watching TV and he is ironing his shirts.
kii_3	juxtaposition	juxtaposition	Paul a-rî-on-a TV a-kî-baac-ag-a n-guo	Paul1.sm-pres-see-fv TV 1.sm-pres.stat?-iron-hab-fv 10-clothes	Paul is watching TV; he is ironing his shirts.
kii_4	îndî	îndî	Susie ni n-dagitaarî îndî ti-mû-toong-û	Susie foc 9-doctor but neg-1sm-rich-fv.stat?	Susie is a doctor but she is poor.

Table 3: A subpart of the example table for Kîîtharaka

Ref	Context	Example	Judgment	Q.	Comment
kii_ev_1	conj-nocontrast-sta	kii_1	felicitous	Q1	
kii_ev_2	conj-nocontrast-epi	kii_2	felicitous	Q1	
...					
kii_ev_24	disj-nspk-sta-exc-2	kii_1	infelicitous	Q2	'na' is not appropriate, corrected to: Susie i ndagitaarî na n'wa wîgue i mwarimû
kii_ev_25	disj-nspk-sta-exc-2	kii_23	felicitous	Q2	

Table 4: A subpart of the evidence table for Kîîtharaka

3. Juxtaposition without any irrealis marking cannot express disjunction (Mauri, 2008);
4. There is no lexicalization for XOR.

To properly test these claims, we need operational criteria for **lexicalization** and **morphological complexity**. For our purposes, we use the **full forms** recorded in the dataset (see Sect. 4—Example table) as the proxy for lexicalization. Note that this is arguably a liberal operationalization of lexicalization. If we do not find any NAND or XOR lexicalization under this liberal definition of lexicalization, we will not find lexicalized NAND or XOR under stricter definitions of lexicalization. We categorize any full form that consists of multiple morphemes as morphologically complex.

The dataset includes a Python script used to test the generalizations (`generalization_checker.py`). Based on the properties of the contexts recorded in the summary table, we have found the following in relation to the four generalizations above. (See the Appendix in §8 for the concrete search criteria.)

1. There is no full form that expresses NAND.
2. There are two full forms that express NOR but do not morphologically contain overt negation.
   - Farsi *na...na...*
   - Greek *ute...ute...*

   There is no occurrence of morphologically simple NOR in the dataset.
3. There is one instance of juxtaposition that expresses disjunction in Igbo (`Igbo_Q2_B2_2`):

   (9) John nwe-re ike bu-ru dokita,
   John have-rV strength cop-rV doctor
   o nwe-re ike bu-ru ode
   he have-rV strength cop-rV write
   akwukwo.
   book
   "John might be a doctor; he might be a writer".

   This example, however, arguably contains a modal element (*nwe-re ike* 'might'; lit. 'have strength') in each disjunct, suggesting that it still conforms to Mauri's (2008) generalization.
4. The following elements express exclusive disjunction according to the judgments collected: Basque *ala*, Dutch *(of) ... of*, Greek *ite ... ite ...*, Kîîtharaka *gati ka ... na ...*, Spanish *o ... o*, Thai *mai gor*, Turkish *ya...ya da...*, Vietnamese *hoặc ... hoặc*

The details of the data in our dataset thus align with at least the first three cross-linguistic generalizations. We will refrain from making broader conclusions about the cross-linguistic generality of

Expression	kboth	kneither	contrast	stative	...	question	fc	can_express	Evidence
juxtaposition	1	0	0	0	...	0	0	1	kii_ev_3,kii_ev_30
na	1	0	0	1	...	0	0	1	kii_ev_1
na	0	0	0	1	...	0	0	0	kii_ev_24
îndî	1	0	1	0	...	0	0	1	kii_ev_5

Table 5: A subpart of the summary table for Kîîtharaka

these hypotheses. Rather, the above discussion highlights the ability of our dataset to support systematic validation of cross-linguistic generalizations and discoveries of their potential counterexamples, which can be fed to more fine-grained language-specific examinations. One such language-specific case study in Thai is detailed in the next section.

5.2. Case study on Thai

In this small case study, we demonstrate how our data collection reveals new empirical insights with potential implications for the cross-linguistic generalizations proposed in previous literature. A notable pattern that emerged in our dataset is that several languages systematically use additive particles (≈ *also, too*) in disjunctive constructions. The additive may combine with a disjunctive connective without much impact on interpretation (e.g. in Turkish), or it may combine with markers of modality to express inclusive disjunction (Igbo, Cantonese). A particularly interesting case of disjunction with an additive particle was found in Thai. As shown in (7) and (8), Thai has a mono-morphemic disjunctive connective (*rŭe*) that is compatible with inclusive and exclusive readings. However, our translation task with the context in (7) revealed an additional disjunctive construction, exemplified in (10), that combines the negation marker *mâi* with the additive particle *gôr* (≈ *also*). Interestingly, while this construction can naturally express the exclusive disjunctive meaning elicited in (7), it is incompatible with the context in (8). In other words, the disjunction in (10) is restricted to exclusive readings.

(10) ✓ exclusive reading (context (7))
inclusive reading (context (8))
Paul pen mŏr **mâi gôr** John pen mŏr.
Paul be doctor NEG ADD John be doctor
"Paul is a doctor or John is a doctor."

In principle, there are two routes to analyzing the construction in (10), both with interesting implications for the typology of disjunction. The first analytical option is a *lexicalization* account, under which *mâi gôr* forms a lexicalized connective with an exclusive disjunctive semantics. Under this account, Thai presents a direct counterexample to the generalization that exclusive disjunction does not lexicalize in natural languages (see Sect. 2.2, 5.1). The second analytical option is a *compositional* account, which derives the disjunctive interpretation from the literal meanings of the negation and the additive. As spelled out in more detail in Mucha et al. (2025), such a compositional account is possible if we assume that a conditional structure underlies the disjunctive interpretation of (10). Specifically, this analysis suggests that the literal meaning of (10) can be paraphrased as *If Paul is a doctor, it is not the case that also John is a doctor*. The compositional analysis straightforwardly derives the incompatibility with inclusive readings — the paraphrased conditional semantically excludes situations in which both Paul and John are doctors — without stipulating an exclusive disjunction meaning for *mâi gôr*. Moreover, this compositional analysis relates Thai to a variety of languages that employ a combination of negation and conditional marking to express disjunctive meaning. Interestingly, however, if the Thai construction in (10) is most accurately analyzed as a type of 'conditional disjunction' in this sense, it still presents a special case. To our knowledge, all cross-linguistically attested disjunctive constructions with conditional structure involve negation of the antecedent of the conditional, and therefore do not semantically restrict the interpretation to exclusive disjunction. A relevant example from the Kwa language Nànáfwɛ (cited from Mauri 2008, 44) is shown in (11).

(11) cɛ́n wjélɛ́ **sɛ́ nán** ánwán jɛ́ ɔ́ tíké
day some if NEG door that it open.PRF
ɔ́ fùndréti jɛ́ ɲín ɔ́n
FOC window that.it slam.PRF FOC
"Sometimes a door opened or a window slammed." (lit. "Sometimes, if it wasn't a door that opened, it was a window that slammed.")

This is in contrast with the Thai construction in (10), where the negation applies to the second clause *John pen mŏr* which corresponds to the consequent in a conditional analysis.

Irrespective of the concrete account, the construction in (10) seems to present a cross-linguistically unpredicted case of disjunction. Under the lexicalization analysis, *mâi gôr* lexicalizes the meaning of exclusive disjunction (XOR), which has previously been precluded in the literature (Katzir and Singh 2013, see also Haspelmath 2007). Under the compositional analysis, (10) exemplifies a

previously unattested strategy of realizing disjunction by combining negation and an additive in a conditional structure.

5.3. Felicity judgments benchmark

Our dataset can also serve as a benchmark for a NLU system's ability to determine the felicity of a connective in context.

As a proof of concept, we explored the extent to which ChatGPT5, one of the leading Large Language Models (LLMs), can provide felicity judgments that agree with our consultants'. Concretely, for each language, we used the prompt in (12).

(12) **Prompt**: I will paste a csv file below that contains two columns: contextDescription and testSentence. For each row, the first column (contextDescription) contains the English description of a context, and the second column(testSentence) contains a sentence in [LAN]. Your task is to determine whether the [LAN] sentence is felicitous in the corresponding context. Output a csv file in plain text. Keep the first two columns exactly the same as the original file. Then record your judgments (felicitous or infelicitous) in a 3rd column named "judgment." And also include a 4th column named "notes" where you can optionally include short explanations for your judgments. Keep the explanations brief and surround them with double quotes.

For certain low-resource languages such as Igbo and Kîîtharaka, ChatGPT5 asked for glosses for some of the connectives. In such cases, it was told to proceed based on its knowledge of the language, with no glosses provided. For Kîîtharaka, it did so after acknowledging it did not have enough competence to provide reliable judgments.

Following Warstadt et al. (2019), we exclude cases marked with ? by consultants, and treat those marked with ??, #, or * as infelicitous. ChatGPT5's agreement rates with the consultant's judgments on positive and negative evidence are shown in Table 6. The agreement rate is generally well above chance level, sometimes even at the ceiling. This suggests that state-of-the-art LLMs exhibit considerable understanding of the truth and felicity conditions of connectives across languages.

Disagreements on positive evidence tend to be due to ChatGPT5's more stringent standard on grammaticality, possibly influenced by prescriptive rules or stylistic preferences. For example, ChatGPT5 judged the following Japanese example involving two occurrences of the disjunctive particle *ka* to be ungrammatical, although it was judged to be perfectly grammatical by the consultant.

(13) Jon-wa ima isya-ka sakka-ka-da
John-TOP now doctor-DISJ writer-DISJ-COP
"John is a doctor or a writer now"

In contrast, disagreements on negative evidence tend to be due to non-truth-conditional/pragmatic aspects of the meaning. For example, given that the two Mandarin Chinese sentences in (15) are both true in context (14), ChatGPT5 judged the juxtaposition to be felicitous, whereas the consultant judged it infelicitous because an additive particle *ye* ("also") is required in the second sentence.

(14) **Context**: Paul works in a hospital. He is not a doctor and he is not a nurse. He works at the reception.

(15) #Baluo bu shi yisheng. Ta bu shi hushi.
Paul NEG is doctor. He NEG is nurse.
"Paul is not a doctor. He is not a nurse"

As another example, in a free-choice context (16), ChatGPT5 judged the Hebrew example (17) felicitous, whereas the consultant judged it infelicitous, commenting that it does not have a free-choice reading and can only mean the speaker is unsure which disjunct is true.

(16) **Context:** There are two sweets on the table. A cupcake and a cookie. Paul can choose either one of them but not both. You don't care which one he chooses.

(17) #O še-Pol yaxol le?exol
CONN comp-Paul can.prs.3sg.m eat.inf
kapkeyk o še-hu yaxol
cupcake CONN comp-he can.prs.3sg.m
le?exol ugiya.
eat.inf cookie
"Paul can eat a cupcake or he can eat a cookie."

6. Conclusions

In this paper, we have described our multilingual dataset of connective semantics. The dataset contains semantic annotations of clausal connectives in a typologically diverse set of 24 languages. The dataset includes systematic information on context-sentence compatibility as evidence for the semantic annotations, including negative judgments (i.e., incompatibility between a context and a sentence involving the target connective). The inclusion of the compatibility data supports detailed examination of cross-linguistic generalizations and their potential counterexamples, as well as benchmarking of sentence felicity judgments by NLU systems.

Akan	Basque	Cantonese	Chinese	Dutch	Farsi	Greek	Hausa
.789 / .559	.710 / .822	.700 / .707	.655 / .897	.789 / .857	.550 / .816	.756 / .981	.824 / .933
Hebrew	**Hindi**	**Hungarian**	**Igbo**	**Japanese**	**Khmer**	**Kiitharaka**	**Korean**
.730 /.708	.739 / .875	.841 / .931	.935 / .917	.471 / 1.00	.769 / 1.00	.897 / .633	.797 / .836
Mapudungun	**Russian**	**Spanish**	**Tagalog**	**Telugu**	**Thai**	**Turkish**	**Vietnamese**
.544 /.757	.870 / .886	.775 / .766	.909 / .792	.739 / .925	.586 / .657	.841 / .862	.973 / .772

Table 6: ChatGPT5's agreement rates with the consultant's judgments on positive/negative evidence

7. Acknowledgements

We deeply appreciate the contributions by our language consultants in building the dataset. We also thank our research assistants, Keyu Dong, James Engels and Fred Whibley, for their assistance during the data collection. This work is supported by the UKRI Future Leaders Fellowship 'Logic in Semantic Universals' (reference: MR/V023438/1).

8. Appendix: Criteria for the evaluation of generalizations

To evaluate the generalizations listed in Sect. 5.1, we have used the following criteria for the relevant connectives defined in terms of the context features in the summary table.

1. A connective expresses **NAND** if it is (i) compatible with a Kp=1 context, (ii) compatible with a Kneither=1 context and (iii) *incompatible* with any Kboth=1 context.
2. A connective expresses **NOR** if it is (i) compatible with a Kneither=1 context, (ii) *incompatible* with any Kboth=1 context, and (iii) *incompatible* with any Kp=1 context.
3. A connective expresses **XOR** if it is (i) compatible only with contexts where the Kboth = 0 and (iii) incompatible with any Kneither=1 context.
4. A juxtaposition involves the connective form $\varnothing$.

9. Ethical considerations and limitations

For data elicitation with native speakers of the target languages, we adhered to The British Association for Applied Linguistics Recommendations on Good Practice and have obtained ethical approval from the Ethics Committee within School of Philosophy, Psychology and Language Sciences, University of Edinburgh, prior to beginning the research.

All native speaker consultants have provided informed consent prior to participation and were able to withdraw from the research without repercussion. The Participant Information Sheet provided participants with full information about what is required in the study, the use and secure storage of data, any risks or benefits to the research, who to contact if they have any concerns about the research, and reminded them of the voluntary nature of the research and their rights to withdraw.

Limitations of our dataset concern representativeness of data in two aspects: the number of native speaker consultants per language and the number of languages. At the time of writing, the dataset only includes judgments from one consultant per language. This limits the representativeness of the judgments across speaker variations. To guarantee the stability and reproducibility of data, we have included metadata files for each language that include information about the speaker's demographic background as well as dialectal information if relevant. We also emphasize that the format of the dataset using context-sentence compatibility allows reproduction of the data with other native speakers.

At the time of writing, the dataset only contains 24 languages, and lacks representations from a number of language families. This limits the conclusions one can draw from the data regarding cross-linguistic generalizations and variation. In the future, we hope to expand the language sample, particularly from regions in the Americas, Oceania and Caucasus.

Finally, our discussion on the use of our dataset as a benchmark for sentence felicity judgments by NLU systems is tentative and exploratory only. We have not compared the results of using different NLU systems (including different LLMs) or using different prompts (our current setup is intended to mirror that for a human consultant). We leave a systematic investigation for future work.

10. Bibliographical References

Moshe E Bar-Lev and Roni Katzir. 2025. Communicative stability and the typology of logical operators. *Linguistic Inquiry*, 56(3):415–437.

Margit Bowler. 2014. Conjunction and disjunction in a language without 'and'. In *Proceedings of Semantics and Linguistic Theory (SALT) 24*, pages 137–155.

Enrico Flor, Nina Haslinger, Hilda Koopman, Eva Rosina, Magdalena Roszkowski, and Viola Schmitt. 2017. Cross-linguistic evidence for a non-distributive lexical meaning of conjunction. In *Proceedings of the Amsterdam Colloquium*, pages 255–264.

Max Glockner, Vered Shwartz, and Yoav Goldberg. 2018. Breaking NLI systems with sentences that require simple lexical inferences. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 650–655, Melbourne, Australia. Association for Computational Linguistics.

Martin Haspelmath. 2007. Coordination. In Timothy Shopen, editor, *Language Typology and Syntactic Description, 2nd ed., vol. 2: Complex constructions*, pages 1–51. Cambridge University Press, Cambridge.

Laurence Robert Horn. 1972. *On the semantic properties of logical operators in English*. University of California, Los Angeles.

Hans Kamp. 1973. Free choice permission. *Proceedings of the Aristotelian Society*, 74:57–74.

Roni Katzir and Raj Singh. 2013. Constraints on the lexicalization of logical operators. *Linguistics and Philosophy*, 36(1):1–29.

Lisa Matthewson. 2004. On the methodology of semantic fieldwork. *International journal of American linguistics*, 70(4):369–415.

Lisa Matthewson. 2011. Methods in cross-linguistic semantics. In Claudia Maienborn, Klaus von Heusinger, and Paul Portner, editors, *Semantics: An international handbook of natural language meaning*, pages 268–284. de Gruyter, Berlin.

Caterina Mauri. 2008. The irreality of alternatives: Towards a typology of disjunction. *Studies in Language*, 32(1):22–55.

Moreno Mitrovic and Uli Sauerland. 2014. Decomposing coordination. In *Proceedings of NELS*, volume 44, pages 39–52.

Anne Mucha, Wataru Uegaki, Mary Amaechi, and Tim Jantarungsee. 2025. Disjunction with additives. In *Proceedings of Sinn und Bedeutung 29*, pages 1071–1087.

Aakanksha Naik, Abhilasha Ravichander, Norman Sadeh, Carolyn Rose, and Graham Neubig. 2018. Stress test evaluation for natural language inference. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*, pages 2340–2353, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Swarnadeep Saha, Yixin Nie, and Mohit Bansal. 2020. ConjNLI: Natural language inference over conjunctive sentences. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8240–8254, Online. Association for Computational Linguistics.

Mario Saltarelli. 1988. *Basque*. Croom Helm, London.

Mark Sammons. 2015. Recognizing textual entailment. *The Handbook of Contemporary Semantic Theory*, pages 523–557.

Benjamin Spector. 2014. Global positive polarity items and obligatory exhaustivity. *Semantics and Pragmatics*, 7:1–61.

Wataru Uegaki. 2023. The informativeness/complexity trade-off in the domain of boolean connectives. *Linguistic Inquiry*, 55(1):174–196.

Alex Warstadt, Amanpreet Singh, and Samuel R. Bowman. 2019. Neural network acceptability judgments. *Transactions of the Association for Computational Linguistics*, 7:625–641.

Yoad Winter. 2002. *Flexibility principles in Boolean semantics: The interpretation of coordination, plurality, and scope in natural language*. MIT press.

Thomas Ede Zimmermann. 2000. Free choice disjunction and epistemic possibility. *Natural Language Semantics*, 8(4):255–290.

11. Language Resource References

Diptesh Das, Thomas Scheffler, Peter Bourgonje, and Manfred Stede. 2018. Constructing a lexicon of English discourse connectives (DiMLex-Eng). In *Proceedings of the SIGDIAL 2018 Conference*, pages 23–32.

Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie, editors. 2005. *The World Atlas of Language Structures*. Oxford University Press, Oxford.

Hilda Koopman and Cristina Guardiano. 2022. Managing data in TerraLing, a large-scale cross-linguistic database of morphological, syntactic, and semantic patterns. In *The Open Handbook of Linguistic Management*, pages 617–630. MIT Press.

Guillaume Marchal, Rudi Scholman, and Volker Demberg. 2021. Semi-automatic discourse annotation in a low-resource language: Developing a connective lexicon for Nigerian Pidgin. In *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*, pages 84–94.

Ponrawee Prasertsom, Apiwat Jaroonpol, and Attapol T. Rutherford. 2024. The Thai Discourse Treebank: Annotating and modeling discourse relations in Thai. *Transactions of the Association for Computational Linguistics*, 12:1–25.

Hedvig Skirgård, Hannah J. Haynie, Damián E. Blasi, Annemarie Verkerk, Sam Passmore, Felix Tiedemann, Christoph Rzymski, Yulia Egorova, Tobias Hofer, Gerhard Jäger, Robert Forkel, Johanna Nichols, Michael Dunn, and Russell D. Gray. 2023. Grambank: A cross-linguistic database of morphosyntactic variation. *Science Advances*, 9(16):eadg6175.

Manfred Stede, Thomas Scheffler, and Ana Mendes. 2019. Connective-Lex: A web-based multilingual lexical resource for connectives. *Discours*, 27:1–35.

Shun Wan, Peter Bourgonje, Heng Xiao, Chih-Wen Ho, and Manfred Stede. 2023. Chinese-DiMLex: A lexicon of Chinese discourse connectives. In *Proceedings of the CODI Workshop at ACL 2023*, pages 110–122.

Derya Zeyrek and Kübra Başıbüyük. 2019. TCL: A lexicon of Turkish discourse connectives. In *Proceedings of the DMR Workshop at ACL 2019*, pages 41–50.