



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Cross-linguistic dataset of force-flavor combinations in modal elements

### Citation for published version:

Uegaki, W, Mucha, A, Hannon, E, Engels, J & Whibley, F 2024, 'Cross-linguistic dataset of force-flavor combinations in modal elements', *Linguistic Variation*.

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Peer reviewed version

### Published In:

Linguistic Variation

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Cross-linguistic dataset of force-flavor combinations in modal elements

{ Wataru Uegaki<sup>◇</sup> Anne Mucha<sup>◇</sup> }  
Ella Hannon<sup>\*</sup> James Engels<sup>◇</sup> Fred Whibley<sup>◇</sup>

<sup>◇</sup>University of Edinburgh    <sup>\*</sup>University of British Columbia

April 8, 2024

## Abstract

We present a cross-linguistic dataset of force-flavor combinations in modal elements, which currently contains information on modal semantics in 24 languages. We discuss theoretical motivations for constructing the dataset, the data collection methodology, as well as the design and the format of the dataset. We also present four case studies using the data: (i) assessment of cross-linguistic generalizations on force/flavor variability; (ii) exploration of generalizations in the lexicalization of negative modality; (iii) investigation of the typology of the morphological encoding of modal strength; and (iv) examination of how future contributes to modality. These case studies illustrate that the dataset supports in-depth assessment of potential cross-linguistic generalizations as well as theory-informed investigations of cross-linguistic variations in modal semantics.

## 1 Introduction

Modality, as expressed by English auxiliaries such as *may* and *must*, is a semantic category for linguistic elements that express notions such as possibility and necessity. Within the linguistic literature, it has been established that modality is a domain with vast cross-linguistic variation (Peterson, 2010; Rullmann et al., 2008; Matthewson, 2010; Bochnak, 2015; Matthewson, 2016; Rullmann & Matthewson, 2018, i.a.). The precise range of this cross-linguistic variation however is a matter of ongoing investigation both in typological and in theoretical linguistics. While there has been considerable progress in cross-linguistic research on modality, a remaining challenge lies in the lack of systematically comparable empirical data that is both descriptively accurate and theoretically informed.

Typological works on linguistic modality provide highly informative, large-scale overviews of modal systems in the languages of the world. However, they often rely on existing grammatical descriptions of the languages under consideration, which can differ widely in precision, background assumptions and terminology, thus impeding detailed comparison of the language-specific modal systems. By contrast, theory-driven cross-linguistic work on modality is usually motivated by highly specific research questions that lead to in-depth investigation of particular aspects of individual modal systems. While extremely fruitful for the development of cross-linguistically valid theories of modality, such works tend to assume a rather narrow perspective on the modal system of the investigated language, and they often aim at testing particular predictions of linguistic theories rather than detecting cross-linguistic patterns.

This paper presents a new dataset on linguistic modality that contributes to closing this gap. It consists of modal questionnaires that contain parallel elicited data on the expression of modal meaning in 24 languages, covering several major language families. The construction of the questionnaires builds on existing elicitation efforts that integrate typological and theoretical perspectives on modal meaning (Vander Klok, 2021) and documents how these languages encode different combinations of modal force and modal flavor (see the next section for more on these notions). The dataset thus provides a resource for systematic comparison of the morphological realization of modal meaning across languages, and a starting point for in-depth investigation of the modal systems of the individual languages included in the database.

The discussion of the dataset is structured as follows. In Sect. 2, we introduce our background assumptions regarding the semantics of modality (2.1) and discuss some existing generalizations on cross-linguistic variation in modal semantics (2.2). We then lay out the motivation for our modal dataset based on these existing cross-linguistic insights (2.3), and briefly compare our database to a related project documented in Guo et al. (2022) (2.4). In Sect. 3, we explain how the data in the dataset were collected, and present the language sample that it covers. Sect. 4 describes the structure of the dataset in more detail, including the concrete content of the questionnaire documents, as well as information on how to access and contribute to the database. In Sect. 5, we present a selection of small case studies, in order to illustrate how the data collected in our dataset may inform pertinent issues in cross-linguistic modal semantics. Sect. 6 concludes the paper.

## 2 Background

### 2.1 Dimensions in the semantics of modals

In this section, we outline our theoretical assumptions about the semantics of modals. These assumptions motivate the kinds of typological and cross-linguistic generalizations in modal semantics we focus on, and provide foundations for the design of our database.

The semantics of modal elements can be understood in terms of variations along at least the following dimensions: *force*, *flavor* and *strength*. Two of these dimensions, modal force and modal flavor, are distinguished in terms of a theoretical framework for modality based on modal logic and possible worlds semantics as developed in Kratzer (1977, 1981, 1991) and summarized in Portner (2009), Hacquard (2011) and Matthewson (2016), among others. Modal force refers to the distinction between necessity ( $\Box$ ) and possibility ( $\Diamond$ ) meanings, which are formalized as universal and existential quantification over possible worlds, respectively. Modal auxiliaries in English, which serve as a reference point for much cross-linguistic work on modality, encode necessity or possibility in their lexical semantics. The contrast is illustrated in (1) with the possibility modal *can* and the necessity modals *must* and *have to*.

- (1) *Context: A violin audition. The conductor says:*
- |  |             |
|--|-------------|
| a. Madeleine <b>can</b> put her violin away now.         | POSSIBILITY |
| b. Madeleine <b>must/has to</b> put her violin away now. | NECESSITY   |
- (Matthewson, 2016, 528)

In addition to possibility and necessity, languages can in principle lexicalize negations of these modal forces: impossibility and non-necessity. For example, the Thai modal *harm* is used to express impossibility while the English *needn't* is used to express non-necessity (Horn, 1989), as

illustrated in the following:<sup>1</sup>

- (2) Yardpubuay **harm** yuu lang hok mornng (Thai)  
 patient.visitor MOD( $\neg\Diamond$ ) exist after six o'clock  
 ‘Visitors can’t stay after 6 o’clock’ IMPOSSIBILITY
- (3) You **needn’t** specify reasons for your decision. NON-NECESSITY

It is worth noting that *needn’t* is morphologically complex and transparently represents negation in its morphological makeup. We will discuss the morphosyntactic complexity of non-necessity items in Sect. 2.2.2.

The term *modal flavor* captures the intuition that modals denote necessity or possibility relative to a restricted set of possible worlds, such as worlds that are compatible with certain rules and regulations (deontic modality), knowledge or evidence (epistemic modality), goals or wishes (teleological/bouletic modality), or someone’s circumstances, dispositions or abilities (circumstantial modality). The flavor of English modals such as *can*, *may* or *must* is not specified in their lexical meaning but depends on the utterance context. In the examples in (4) (adapted from Hacquard 2011, 1485), we illustrate some crucial modal flavor distinctions by means of modification with ‘in view of’-phrases, which make the modal flavor explicit:

- (4) a. **Epistemic**  
 (In view of the available evidence,) John *may / must* be the murderer.
- b. **Deontic**  
 (In view of his parents’ orders,) John *may* watch TV, but he *must* go to bed at 8 pm.
- c. **Circumstantial/ ability**  
 (In view of his physical abilities,) John *can* lift 200 lbs.
- d. **Teleological**  
 (In view of his goal to get a PhD,) John *must* write a dissertation.
- e. **Bouletic**  
 (In view of his desire to retire at age 50,) John *should* work hard now.

Previous semantic research has discovered interesting variation in modal systems across languages. For instance, in some languages, such as St’át’imcets, Gitksan and Tlingit, modals exhibit variable, context-dependent modal force but lexicalize modal flavor (e.g. Rullmann et al., 2008; Peterson, 2010; Cable, 2017). The following examples from Rullmann et al. (2008) show that the St’át’imcets modals *k’a* and *ka* specifically encode epistemic and deontic flavors respectively,<sup>2</sup> although they are compatible with both possibility and necessity readings:

- (5) **Variable-force epistemic modal *k’a* in St’át’imcets** (Rullmann et al., 2008, 320–321)

<sup>1</sup>Glossing abbreviations: ACC = accusative case; AV = actor voice; AUX = auxiliary; CIRC.POS = circumstantial possibility modal; COP = copula; DAT = dative case; DECL = declarative marker; DEF = definiteness marker; DET = determiner; DEIC = deictic element; DIR = directive transitivizer; FUT = future; IMPF = imperfective; IND = independent mood; INF = infinitive; LOC = locative; PAST = past; PL = plural; POSS = possessive; PRT = particle; Q = question marker; REFL = reflexive; REL = relative clause marker; SG = singular; SM = subject marker; SR = switch reference; SUBJ = subjunctive; TAM = tense, aspect, modality NOM = nominative case; MOD = modal element; NEG = negation YNQ = yes/no-question marker.

<sup>2</sup>The item *ka* also has an irrealis use, omitted here for expository purposes.

- a. Wa7 **k'a** qwenúxw.  
 IMPF MOD sick  
 'He may/must be sick.'
- b. t'ak **k'a** tu7 kents7á ku míxalh  
 go.along MOD then DEIC DET bear  
 'A bear must have gone by around here.'
- c. *Context: His car isn't there.*  
 plan **k'a** qwatsáts  
 already MOD leave  
 'Maybe he's already gone.'
- (6) **Variable-force deontic modal *ka* in St'át'imcets** (Rullmann et al., 2008, 328)
- a. kan **ka** kw-en-s ulhcw  
 YNQ MOD DET-1SG.POSS-NOM enter  
 'Should/can/may I come in?'
- b. lán-lhkacw **ka** áts'x-en ti kwtámts-sw-a  
 already-2SG.SUBJ MOD see-DIR DET husband-2SG.POSS-DET  
 'You must/can/may see your husband now.'

Necessity modals can be distinguished according to yet another dimension: *modal strength*. The distinction can be illustrated by the contrast between the strong necessity modals *must/have to* and the weak necessity modals *should/ought* in English. Under their deontic readings, for instance, *must* places a stronger requirement than *should/ought*, as is evident in the following example:

- (7) Employees must wash their hands. Non-employees really ought to wash their hands, too.  
 (von Fintel & Iatridou 2008, 115<sup>3</sup>)

More precisely, weak necessity, strong necessity and possibility can be distinguished based on the following two criteria from Rubinstein (2021). Firstly, strong and weak necessities are alike in disallowing conjunction of mutually exclusive propositions, unlike possibilities which allow such conjunctions:

- (8) a. #It must be raining and it must not be raining. STRONG NECESSITY  
 b. #It should be raining and it should not be raining. WEAK NECESSITY  
 c. It may be raining and it may not be raining. POSSIBILITY

Secondly, among necessity modals, strong necessity entails weak necessity, but not vice versa. This entailment relation gives rise to a scalar implicature which can be overtly reinforced, cancelled, or suspended (Silk, 2022; Weingartz & Hohaus, to appear):

- (9) I ought to help the poor...  
 a. In fact, I must. CANCELLATION  
 b. But I don't have to. REINFORCEMENT  
 c. Maybe I have to. SUSPENSION

To our knowledge, weak *possibility* modals have not been reported in the literature, although they may in principle exist at least under certain theoretical analyses of the strength distinction

<sup>3</sup>From a cartoon appearing in the July 31, 2006 issue of the *New Yorker*

(see Rubinstein, 2012; Vander Klok & Hohaus, 2020 for discussion). Thus, for the purpose of this paper and in our database, we assume that the strength distinction is only applicable to necessity. This gives rise to the categorization of modal meanings as illustrated in Table 1.

		Epistemic	Deontic	Teleological	Circumstantial	Bouletic
Necessity	Strong	<i>must</i>	<i>must</i>	<i>have to</i>	<i>have to</i>	<i>want</i>
	Weak	<i>should</i>	<i>ought/should</i>	<i>should</i>	??	??
Possibility		<i>may/might</i>	<i>may</i>	<i>can</i>	<i>can</i>	—
Impossibility		<i>can't</i>	<i>can't/mustn't</i>	<i>can't</i>	<i>can't</i>	—
Non-necessity		??	<i>needn't</i>	<i>needn't</i>	??	??

Table 1: The categorization of modals based on force (the vertical axis: necessity/possibility/impossibility/non-necessity), flavor (the horizontal axis), and strength distinctions (the vertical axis: strong vs. weak). We assume the strength distinction only within necessity. The table also contains examples of English modal expressions in each cell, but this is only for illustrative purposes.

Since it will be relevant for some of our case studies in Sect. 5, let us also briefly introduce the temporal dimension of modal interpretation. We follow Condoravdi (2002) in assuming a distinction between *temporal perspective* (TP) and *temporal orientation* (TO). In a nutshell, a modal’s temporal perspective corresponds to the time at which the evidence for the use of the modal is evaluated. The temporal orientation of a modal is the time at which the modalized situation is temporally located. Temporal orientation in particular can be shown to correlate with the aspectual interpretation of the modalized situation. This is illustrated by the English examples in (10)–(12), all of which involve the epistemic possibility modal *might*. While all the target sentences in (10)–(12) are naturally interpreted with a present temporal perspective (i.e., the relevant evidence for the use of the epistemic statement is obtained at the utterance time), their temporal orientation differs. The target sentence in (10), containing the aspectually unmarked event predicate ‘cry’, conveys that the possible crying event is located in the (possibly very near) future. With a stative predicate or a progressive-marked event, as in (11), *might* expresses a conjecture about a present situation, i.e. a present-oriented epistemic statement. In (12), perfect aspect marking gives rise to a past oriented reading, i.e. the crying took place before the (present) perspective time of the modal.

- (10) *Context: You are meeting your friend with her baby, who looks unhappy. Your friend tells you that the baby is a bit shy with new people, and he says:*  
 She might cry. FUTURE TO
- (11) *Context: You are meeting your friend at home, and his baby is next door with the neighbors. Your friend is nervous and wants to go and check on the baby. He says:*  
 a. She might be crying. PRESENT TO  
 b. She might be scared. PRESENT TO
- (12) *Context: You are meeting your friend with his baby. The baby’s cheeks are wet, and your friend says:*  
 She might have cried. PAST TO

Languages vary in how transparently they encode temporal perspective and temporal orien-

tation, which partly depends on their tense/aspect systems as well as on the syntactic properties of the modals. As shown above, modal future orientation arises with aspectually unmarked event predicates in English, while present and past orientation can be enforced by progressive and perfect marking, respectively. However, it has been argued that some languages whose temporal system includes overt prospective aspect markers transparently indicate future orientation as well. This is illustrated by the example in (13) from Matthewson’s (2013) work on Gitksan. The sentence contains the future/prospective morpheme *dim*, which in Gitksan overtly encodes, and in fact forces, future orientation of modals.

- (13) yugw=**imaa/ima**'=hl **dim** wis (Matthewson, 2013, 365)  
 IMPF=**epis**=CN FUT rain  
 ≠ “It might have rained.” / ≠ “It might be raining.” / “It might rain (in the future).”  
 # Context: You see puddles, and the flowers looking fresh and damp. (PAST TO)  
 # Context: You hear pattering on the roof. (PRESENT TO)  
 ✓ Context: You hear thunder, so you think it might rain soon. (FUTURE TO)

For illustration of a contrast in temporal perspective, consider the English sentences in (14) and (15). They contain *have to*, which inflects for tense. The sentence in (14) conveys that present circumstances are such that it is necessary for Paula to take the car in order to reach a goal/destination (teleological flavor) or fulfil an obligation (deontic flavor). The example in (15), with past tense marking on the modal, shows a reading of *have to* under which, at a specified past time (‘last week’), it was necessary for Paula to take the train to reach her destination. Crucially, the circumstances giving rise to that necessity obtain at the sentence’s past reference time.

- (14) Paula has to take the car. PRESENT TP  
 (15) *Context: Last week, Paula had an appointment in Glasgow. There was a public transport strike and all the trains and buses got canceled.*  
 Paula had to take the car. PAST TP

Scholars have investigated the typology of modal elements based on the type of data exemplified above, and posited several potential cross-linguistic generalizations as well as points of variation. In the following we will survey three active domains of research in the typological variation and generalization in modal force, modal strength and modal flavor.

## 2.2 Existing generalizations and typology

### 2.2.1 Generalizations on force/flavor variability

Nauze (2008) proposes that natural language modals can be underspecified either in the force or in the flavor dimension, but not in both. The generalization can be stated as follows:

- (16) **Nauze’s (2008) Generalization**  
 Modal elements can have multiple meanings only along a unique axis out of the force and flavor axes: they either vary with respect to force or they vary with respect to flavor, but they cannot vary on both axes.

(adapted from Nauze 2008, 222)

The generalization rules in the English-type system, where modals have multiple meanings only

along the flavor dimension, as well as the St’át’imcets-type system, where modals have multiple meanings only along the force dimension. In contrast, it rules out a language where a modal element can express multiple forces and multiple flavors at the same time.<sup>4</sup>

A counterexample to Nauze’s generalization has been found in the modal system of the Washo language (Bochnak, 2015). As exemplified below, the modal verb *é* in Washo is underspecified with respect to both force and flavor.

(17) **Variable-force and variable-flavor modal *é* in Washo** (Bochnak, 2015)

- |    |   |                       |
|----|---|-----------------------|
| a. | dé eš- áṅaw-i-š yéweš gum-béyc’ig-i-gi k’-é -i<br>snow-good-IND-SR road REFL-close-IND-REL 3-MOD-IND<br>‘It’s snowing a lot, so the road must be closed.’ | EPISTEMIC NECESSITY   |
| b. | bévali k’-é -hel-i-gi k’-é -i<br>Beverly 3-be-SUBJ-IND-REL 3-MOD-IND<br>‘It might be Beverly.’  | EPISTEMIC POSSIBILITY |
| c. | súku baṅáya -é -i-š-gi k’-é -i<br>dog outside 3-be-IND-SR-REL 3-MOD-IND<br>‘The dog has to stay outside.’   | DEONTIC NECESSITY     |
| d. | wádiṅ hé:š um-p’áyt’i -giš-uwe k’-é -i<br>now Q 2-play-along-hence 3-MOD-IND<br>‘Now are you allowed to come play?’                                       | DEONTIC POSSIBILITY   |

In an effort to refine the cross-linguistic generalization along the lines of (16), Steinert-Threlkeld et al. (2023) have recently proposed a new generalization which they call INDEPENDENCE OF FORCE AND FLAVOR, formulated as follows:

(18) **Independence of force and flavor** (Steinert-Threlkeld et al., 2023)

All modals in natural language satisfy the independence of force and flavor property: if a modal can express the pairs  $(fo_1, fl_1)$  and  $(fo_2, fl_2)$ , then it can also express  $(fo_1, fl_2)$  and  $(fo_2, fl_1)$ .

In other words, the generalization states that a modal element must be able to express all possible combinations of forces and flavors it is compatible with. Note that the generalization rules in the English-type modals that only vary in flavors and the St’át’imcets-type modals that only vary in force. In addition, it allows the Washo modal *é* discussed above as it can express all four combinations of necessity/possibility and epistemic/deontic flavors. A counterexample to (18) would, for example, be an item that expresses deontic necessity and epistemic possibility, but not deontic possibility.

### 2.2.2 Cross-linguistic generalizations in negative modality

Another existing line of research into the cross-linguistic generalization in modal semantics concerns negative modality. In particular, Horn (1989) suggests that non-necessity is rarely lexicalized in natural language. This generalization follows the general constraint against lexicalization of the Aristotelian ‘O-corner’ operator documented in other functional domains such as connectives (lack of \*NAND) and quantifiers (lack of \*NALL) (Horn, 1972, 1989). Horn notes, though, that the constraint is relatively weak in the modal domain in comparison to those in

<sup>4</sup>Vander Kloek (2013) proposes a slightly modified generalization, where a modal system allows variability along one dimension within an epistemic/non-epistemic flavor category.



the quantificational and connective domains. Whereas O-corner quantifiers/connectives ‘never lexicalize’ there are plausible lexicalizations of a non-necessity modal auxiliary such as *needn’t* in English (Horn, 1989, 260). There are also lexical items that express non-necessity in adjectival and nominal categories, such as *unnecessary* and *uncertainty*. Furthermore, Kuhn & Pasalskaya (2022) have recently noted that French Sign Language (LSF) contains a lexicalized modal auxiliary for non-necessity.

Given these counterexamples, it would be too strong to state that non-necessity never lexicalizes. Still, available evidence seems to support the following implicational generalization stated in terms of morpho-syntactic complexity (Horn, 1972; Kuhn & Pasalskaya, 2022).<sup>5</sup>

(19) **Relative complexity of the O-corner modality**

An expression for the O-corner modality is at least as morpho-syntactically complex as modal expressions for the other three corners, i.e., possibility, necessity, and impossibility.

English examples mentioned above, *needn’t*, *uncertainty*, and *unnecessary*, are all morphologically complex, and thus are at least as complex as modal expressions for possibility, necessity, and impossibility in English. Also, the LSF non-necessity modal is not simpler than the other LSF modal auxiliaries. On the other hand, a counterexample to this generalization would, for example, involve a mono-morphemic non-necessity modal and morphologically complex possibility and necessity modals.<sup>6</sup>

### 2.2.3 Typology of the encoding of modal strength

The existing literature on cross-linguistic modal semantics does not only concern potential *generalizations* across languages, but also systematic patterns of *variation*. In addition to allowing assessment of cross-linguistic generalizations, the database we will present in this paper supports investigation into empirical questions about cross-linguistic variation in modal semantics, e.g., how different variants are attested across languages and whether there are new variants not documented in the literature.

A growing body of literature has investigated the cross-linguistic variation in the encoding of modal strength, which the current database is well-suited to examine. In this domain, at least four types of morpho-syntactic encoding of the strong/weak necessity distinction have been identified cross-linguistically:

**A. Lexical distinction** The distinction between strong and weak necessity can be made as a lexical distinction, as in the case of *must* vs. *should/ought* in English.

**B. Strong necessity + additional morphology** In a number of languages, weak necessity is expressed by combining a strong necessity modal with additional morphology. The nature of this additional morphology depends on the language. Von Fintel & Iatridou (2023)

---

<sup>5</sup>Here, the scale of morphosyntactic complexity can be roughly defined as follows:

- (i) Morphologically simplex word < Morphologically complex word < Syntactically complex constituent < Non-constituent (Kuhn & Pasalskaya, 2022, 2)

<sup>6</sup>In this paper, we will not delve into the *explanations* of cross-linguistic generalizations and stay at the level of description of empirical generalizations. Thus, here, we simply mention that several explanations for the lack of lexicalization in the domain of connectives have been proposed in the literature, including those in terms of principles of lexicalization in terms of Gricean reasoning (Horn, 1989; Katzir & Singh, 2013), informativeness-complexity tradeoff (Uegaki, 2022) and the notion of communicative stability (Bar-Lev & Katzir, 2023).

have identified a wide-spread pattern in European languages in which a morphological marking employed in counterfactual conditionals is added to a strong necessity modal to produce weak necessity. This can be seen in the following Greek examples from von Fintel & Iatridou (2023):

- (20) **Tha eprepe** na plinis ta piata ala dhen ise ipexreomenos na to kanis  
 FUT must.PAST NA wash the dishes but NEG are obliged NA it do  
 ‘You ought to do the dishes but you are not obliged to do it.’
- (21) a. An efevge simera **tha efane** tin ali evdhomadha  
 if left today FUT arrive.PAST.IMP the other week  
 ‘If he left today, he would arrive next week.’  
 b. An ton iche xtipisi to aftokinito **tha iche** pethani  
 if him had hit the car FUT have.PAST died  
 ‘If the car had hit him he would have died.’

In (20), weak necessity is expressed by combining the strong necessity modal *prepi* with the future marker *tha* and the past morphology. Notably, the same marking is observed in the consequent of counterfactual conditionals, as one can see in (21).

In Javanese, weak necessity is expressed by a combination of strong necessity and a designated morpheme (Vander Klok & Hohaus, 2020, 2)(2):

- (22) Wong wong jawa **kudu-ne** iso ngomong kromo, terus anak-e  
 person person java must-NE CIRC.POS AV.talk high.speech then child-DEF  
 rojo yo **kudu** iso.  
 king PRT.yes must CIRC.POS  
 ‘Javanese people ought to be able to speak Krama, and the Sultan’s son has to be able to.’

Here, one can see that the weak necessity is expressed by the combination of the strong necessity modal *kudu* and the additional morpheme *ne*, which Vander Klok & Hohaus (2020) argue to be a designated morpheme not utilized e.g., in counterfactual conditionals. Yet another pattern is observed in Logoori, where an anticausative morpheme is attached to necessity for weak necessity (Gluckman & Bowler, 2020).

**C. Comparative paraphrase** Rubinstein (2014) notes that Hebrew utilizes a comparative paraphrase to express weak necessity:

- (23) **Yoter tov** še-hu yitpater, aval hu lo xayav lehitpater.  
 more good that-he will.resign but he NEG must resign  
 ‘It is better that he resign but he doesn’t have to resign.’  
 (Rubinstein 2014, p. 526, (21a))

**D. Unmarked distinction** Finally, Weingartz & Hohaus (to appear) argue that some languages have a morphologically unmarked distinction between strong and weak necessity. In the following examples from Afrikaans and Samoan, the first and the second clause involve the same modal element, although they express strong and weak necessity respectively.

- (24) Werkers **moet** hande was. Nie-werkers **moet** ook hulle hande was. (Afrikaans)  
workers MOD hands wash not-workers MOD also their hands wash  
‘Employees must wash hands. Non-employees should also wash their hands.’
- (25) Mo tagata fai=galuega: E **tatau** ona fufulu mamā lima. Mo le mamalu  
for people make=work TAM MOD that wash.PL clean hand for the dignity  
lautele: E matuā **tatau** foi ona fufulu mamā lima. (Samoan)  
general TAM really MOD also that wash.PL clean hand  
‘For employees: You must wash hands. For the general public: You should really  
also wash hands.’

An open issue in the literature is what constitutes the grammatical, typological and areal factors that determine the various strategies for encoding the strength distinction.

### 2.3 Motivation for a cross-linguistic database

Informed observations and hypotheses of the kind discussed in the previous subsections can be exposed to empirical scrutiny to test their universal status. Cross-linguistic semantic studies on modality over the years have accumulated a wealth of data that allow us to evaluate the empirical status of various generalizations in modality. To complement existing data and to further the assessment of cross-linguistic generalizations, we present a dataset that contains fine-grained information about the semantics of modal expressions in 24 languages.

The current dataset is designed to address several challenges associated with the evaluation of generalizations based on existing sources. First, since existing sources may use different methods and contexts to elicit modal expressions, cross-linguistic comparison is not always straightforward. Terminological discrepancies for modal flavors across different sources present further challenges (see Portner 2009, 138-141 and Nuyts 2006 for relevant discussion). Since the data in the dataset are collected based on a uniform methodology and terminology (to be discussed in Sect. 3), they are well-suited to support consistent comparison across languages.

Moreover, data in some existing sources predominantly constitute positive evidence, i.e., data showing that certain modal expressions can express a certain force/flavor combination. To properly assess some of the generalizations discussed in the previous subsections, however, we also need negative evidence, i.e., data showing that certain modal expressions cannot express a certain force/flavor combination.

Although the number of sample languages currently available in the dataset is still small and is by no means representative of the world’s languages, our data offers a first step towards several cross-linguistic research in modal semantics, as shown in the case studies to be discussed in Sect. 5. Furthermore, the database and the data collection methodology are designed to make it easy for researchers to contribute data in new languages to the dataset and expand its empirical scope.

### 2.4 An existing database: Guo et al. (2022)

In a recent paper, Guo et al. (2022) offer a cross-linguistic database of modal semantics which supports evaluation of typological generalizations. Since the rationale behind the creation of Guo et al.’s database is close to ours, it is important to discuss the difference between Guo et al. (2022)’s database and the current dataset. However, since the data in Guo et al. are mostly

based on existing sources on individual languages, the database inherits the issues regarding divergent data collection methods and terminological discrepancies. Furthermore, Guo et al.’s data only include information about whether certain modal expressions can or cannot express specific force/flavor combinations. That is, the database lacks concrete example sentences and contexts that illustrate the compatibility/incompatibility between modal elements and force/flavor combinations. Such fine-grained qualitative information is important in extracting theoretically-informed conclusions from the data. One case in point concerns the phenomenon of polarity-sensitive variable force modality (Deal, 2011; Jeretič, 2021). A modal item may seem compatible with both necessity and possibility forces at first sight, but closer examination of the examples may reveal that the force is conditional on the polarity of the syntactic environment in which the modal occurs. In this type of situation, further investigation is required to establish the underlying force of the modal element, as it is possible that the apparent variable-force behavior can be explained by the interaction of a non-variable force and additional semantic/pragmatic operations (which are constrained by polarity). Such an investigation, however, is impossible if the available data only tell us whether an item is compatible with multiple forces, without telling us the syntactic environments in which the different forces show up. Additionally, the grammatical properties of the relevant modal items, such as their morpho-syntactic makeup, are not included in Guo et al.’s database. This omission can also lead to missing typological observations that refer to the syntactic properties of a modal item, such as those relevant for the encoding of modal strength as introduced in Sect. 2.2.3.

Since our dataset contains concrete examples and information about the grammatical properties and morpho-syntactic makeup of the relevant modal items, it supports the fine-grained investigation which would be impossible with Guo et al.’s database.<sup>7</sup> In Sect. 5.4 and Sect. 5.3, we give brief case studies of such investigations using our dataset. Specifically, in Sect. 5.4, we consider how future marking contributes to modal semantics in a variety of languages, sometimes through morpho-syntactic combination with other elements in the modal inventory. In Sect. 5.3, we consider how our data fit the existing typology of the encoding of modal strength, with reference to the grammatical properties of the morphemes that make up weak necessity expressions.

This said, our intention is not to *replace* Guo et al.. Rather, we see the two databases to be complementing each other. The type of quantitative machine-readable data in Guo et al.’s database is crucial for making statistical inferences based on the data and comparing natural language data with modelling results (Imel et al., 2023). For this reason, the data available in our database have also been converted to the quantitative format and have been incorporated into Guo et al.’s database. Crucially, though, the fine-grained qualitative data cannot be retained in the conversion process and are only available in our database.

---

<sup>7</sup>This said, it should also be acknowledged that our data collection methodology is limited in terms of the range of grammatical features and contexts it systematically tests. For example, with respect to polarity sensitivity, we have only tested negation as a polarity-reversing environment, and have not systematically investigated the effect of other relevant environments, such as the antecedent of conditionals, the restrictor of “every”, and questions. Thus, although our data will be a basis for a systematic investigation of the polarity-sensitivity of modal elements, they are still not enough for a full examination.

### 3 Methodology of data collection

#### 3.1 The modal questionnaire

For the elicitation of the data in our database, we used an adapted version of the revised modal questionnaire for cross-linguistic use provided by Vander Klok (2021). This questionnaire is designed specifically to elicit different combinations of modal force/strength and modal flavor, and to facilitate systematic comparison across languages, which makes it particularly suitable for detecting and investigating cross-linguistic patterns in modal semantics. Moreover, the questionnaire is designed such that it can be employed for different elicitation methods, including translation in context tasks, acceptability judgment tasks or forced choice tasks (for discussion of these elicitation techniques see e.g. Matthewson 2004, 2011).

Let us illustrate the workings of the questionnaire with some example contexts that vary with respect to modal flavor, modal strength and modal force. The context in (26) sets up a situation in which rules and regulations for hospital visits are evoked, i.e., a deontic conversational background. The target sentence, which contains a necessity modal compatible with this context, is marked in bold. The researcher can choose to present the target sentence in the context language as shown below and ask the consultant to translate it into the target language. Alternatively, the researcher may construct a sentence in the target language and ask the consultant to judge whether the sentence could be felicitously uttered in the relevant context, i.e. with a deontic necessity reading.

- (26) *Context (deontic necessity)*: You are going to visit your friend in the hospital. When you enter into the hospital, you stop at the information desk to inquire what room your friend is in. But the woman at the information desk tells you that you can't visit your friend now because it's already 8pm. She says, "I'm sorry, the hospital regulations say that... **Visitors must leave by 6pm.**"

The context in (27) elicits necessity as well, but in a different flavor. It describes relevant background knowledge about John's usual whereabouts, thereby establishing an epistemic conversational background. Since English modals lexically encode force rather than flavor, the English target sentence contains *must* in both (26) and (27). As discussed in Sect. 2.1, however, this is not expected to be the case in all languages: in a language like St'át'imcets, the contexts in (26) and (27) would elicit different modals.

- (27) *Context (epistemic necessity)*: You know that John goes to school at 9am every day. You look at the clock and see it is 9am, therefore... **John must be at school.**

Compare this to the context in (28), which keeps the (deontic) flavor constant when compared to (26), but varies the modal strength. In other words, (28) elicits deontic weak necessity. In contrast to the context in (26), (28) specifies a rule-based ideal rather than a strict requirement – a meaning that English expresses with modals such as *ought* or *should*.

- (28) *Context (deontic weak necessity)*: In England, it is recommended that face coverings be worn in stores, but it is not a legal requirement. You plan on going shopping, and you think to yourself ... **I ought to wear a face covering.**

Finally, a context differing from (26) in modal force is given in (29). It specifies permission against the background of a teacher's authority over the rules that her students have to follow.

Note also that this context explicitly excludes a circumstantial flavor reading by specifying that Beth does not have the ability to swim.

- (29) *Context (deontic possibility)*: Beth’s teacher told her class that it was okay to go swimming, but Beth doesn’t want to because she has never learnt to swim! However, as far as the teacher’s directions are concerned ... **Beth can go swimming.**

For our cross-linguistic elicitations we complemented Vander Klok’s (2021) questionnaire by adding contexts for eliciting expressions of negative modality, i.e. *non-necessity* ( $\neg\Box p$ ) and *impossibility* ( $\neg\Diamond p$ ). Like in the positive cases illustrated above, we presented our consultants with contexts that specify epistemic, deontic, teleological, or pure circumstantial conversational backgrounds. In many cases, we used the same or very similar context descriptions as adapted from Vander Klok (2021) and simply changed the target sentences. This can be seen in (30) below. The same context description as in (26), originally constructed to elicit deontic necessity meaning, is equally suitable to elicit the deontic impossibility meaning that English would express with *mustn’t* or *can’t*, i.e. with the combination of a modal auxiliary and morphological negation.

- (30) *Context (deontic impossibility)*: You are going to visit your friend in the hospital. When you enter into the hospital, you stop at the information desk to inquire what room your friend is in. But the woman at the information desk tells you that you can’t visit your friend now because it’s already 8pm. She says, “I’m sorry, the hospital regulations say that ... **Visitors mustn’t stay after 6pm.**”

Interestingly, not all languages express impossibility by transparently combining negation and a possibility or necessity modal as in (30). In our cross-linguistic study, we identified several modal systems that contain lexicalized impossibility modals. We will discuss this in more detail in the case study in Sect. 5.2.

For completeness, let us also illustrate the elicitation of non-necessity with the example in (31). Again, this example adapts a context created for eliciting positive deontic modality, in this case the weak necessity context in (28), and modifies the target sentence to express non-necessity.

- (31) *Context (deontic non-necessity)*: In England, it is recommended that face coverings be worn in stores, but it is not a legal requirement. You plan on going shopping, and you think to yourself ... I ought to wear a face covering, but **I don’t have to / needn’t wear a face covering.**

As noted above, the questionnaire systematically crosses the force/strength distinctions between (strong) necessity, weak necessity and possibility with deontic, epistemic, teleological and pure circumstantial modal flavors.<sup>8</sup> Each combination of force/strength and flavor was tested with at least two and up to five examples, some of which specifically contrast the intended meaning with a closely related one (e.g. deontic possibility as opposed to circumstantial possibility in example (29)) or vary interpretational properties that may influence the range of expressions available to express certain modal meanings (e.g. the predicate type, temporal orientation, or properties of the sentential subject). The complete questionnaire thus comprises over 50 modal contexts for elicitation. Hence, while in some languages we did not elicit the whole range of

---

<sup>8</sup>In addition, bouletic necessity contexts were used for eliciting modal meanings along the lines of English “want”.

examples in cases where no variation between similar contexts occurred, our elicitations yield a fairly comprehensive overview of the modal systems in the languages we investigated.

### 3.2 Language sample and elicitation procedure

In (32), we present the full list of languages for which we elicited data on modality by use of this adapted modal questionnaire, grouped by language family where applicable. The language sample was chosen to strike a balance between typological diversity and feasibility: on the one hand, we aimed to cover as many major language families as possible, as well as the main sub-groups of the Indo-European family. On the other hand, we focused on languages that were accessible to us via existing contacts to native speakers. For this reason, the sample contains a relatively large number of European languages as well as languages with institutional and stable, non-endangered status.

- |      |    |  |                  |
|------|----|--|------------------|
| (32) | a. | Dutch, Farsi, Greek, Hindi, Russian, Spanish | (Indo-European)  |
|      | b. | Vietnamese, Khmer                            | (Austro-Asiatic) |
|      | c. | Thai   | (Kra-Dai)        |
|      | d. | Cantonese, Mandarin                          | (Sino-Tibetan)   |
|      | e. | Hausa, Hebrew                                | (Afro-Asiatic)   |
|      | f. | Hungarian                                    | (Uralic)         |
|      | g. | Akan, Igbo, Kĩĩtharaka                       | (Niger-Congo)    |
|      | h. | Tagalog                                      | (Austronesian)   |
|      | i. | Turkish                                      | (Turkic)         |
|      | j. | Telugu                                       | (Dravidian)      |
|      | k. | Basque                                       |                  |
|      | l. | Japanese                                     |                  |
|      | m. | Korean                                       |                  |
|      | n. | Mapudungun                                   |                  |

The data elicitation was conducted in online interviews with our language consultants, via video conference calls. With the exception of the language consultants for Khmer and Mapudungun, all native speakers we worked with have a background in linguistics, but with varying levels of relevant expertise: some consultants were working linguists specialized in semantics, while others had no formal training in semantic theory. The elicitations were structured into 90 minute sessions. On average, the completion of the modal questionnaire took three such sessions, although the expenditure of time differed depending on factors such as the researchers' familiarity with the language, the complexity of the elicited data and the extent of follow-up elicitations (e.g. acceptability judgment tasks). As compensation for their time and effort, the consultants received a payment of 9 GBP per hour of work.

Our cross-linguistic data collection focused on translation in context tasks. Consultants were presented with the linguistic context description in English, along with a target sentence that expresses the intended meaning (see the examples in Sect. 3.1). The consultants were then asked to provide a translation of the target sentence in their native language that is felicitous in the given context. Presenting target sentences in context, rather than in isolation, constitutes one of several measures taken to avoid potential pitfalls of translation tasks in linguistic fieldwork. As already discussed by Matthewson (2004), adding discourse contexts for translation tasks is particularly important when the target expressions are potentially ambiguous or context-sensitive. The presented contexts clarify, on an intuitive level, which interpretation is intended for any

particular example, and thus take account of the fact that many natural language modals show variability in some of the semantic dimensions discussed in Sect. 2 (e.g., variable modal flavor in the English translation prompts). Another potential worry relating to the translation task is that the structure of the English target sentence might influence the consultant to produce a similar structure, instead of a potentially more natural construction, in their native language. However, in our study consultants always had the possibility to provide several translations, and to comment on their preferences and intuitions on subjective differences between the provided translations. These translation data provide (positive) evidence that the elicited construction can convey the intended modal meaning. As already mentioned in Sect. 2.3, additional negative evidence is necessary when the goal is to model the meaning of a targeted expression. Accordingly, whenever we were interested in semantic generalizations regarding specific modal elements, we complemented the translation tasks with consultants’ judgments on acceptability and felicity in context, as will be illustrated in the following sections.<sup>9</sup>

In practical terms, our elicitations were conducted jointly by one investigator and one research assistant, and the questionnaires were shared with the consultant in a collaborative online document for the duration of the elicitation sessions. In a typical data collection meeting, one of the researchers read the context and the target sentence to the consultant in English<sup>10</sup>, and the consultant wrote down the appropriate version of the target in their native language. The target sentences were glossed on the spot with the help of the consultant unless one of the researchers had sufficient knowledge of the target language to gloss the sentences afterwards. Any relevant comments by the consultants were documented, along with volunteered examples, ad hoc judgments etc. We noted these comments directly in the questionnaire and edited the questionnaire for readability and completeness after each session. Final editing in preparation for publication of the questionnaires took place after data elicitation was completed.

Where possible, the complete and edited questionnaires were reviewed by a native speaker linguist other than the primary language consultant, with the reviewer’s comments taken into consideration in the final version of the questionnaire.

## 4 Structure and format of the database

The database consists of documentations of the data collection—which we call questionnaire documents—for each language. In this section, we discuss the content of the questionnaire documents as well as how other researchers can access the dataset and contribute new data to it.

### 4.1 Content of the questionnaire documents

The questionnaire documents primarily consist of the following information, which we will detail with examples below:

1. Metadata about the language and the native speaker consultant.

---

<sup>9</sup>In this paper, we cannot elaborate on the relation between translation tasks and judgment tasks in semantic fieldwork. We refer the reader to Matthewson (2004); Deal (2015); Sardinha (2022) for relevant discussion from different perspectives.

<sup>10</sup>An exception to this was the data elicitation in Mapudungun, where Spanish was used as the language of communication.



2. Translation of the target sentences (embedded within appropriate contexts) into the target language.
3. A table of elicited modal expressions organized in terms of their force, flavor and strength.

**Metadata** The metadata consists of the language name, the ISO 639-3 code and the Glottolog code, the relevant demographics of the native speaker consultant who provided the data, as well as the elicitation dates. As the demographic information of the consultant, we include the gender as well the dialect of the language based on their self report. Where relevant, we also include other languages the consultant is proficient in, in order to track the potential effect of multilingualism. Figure 1 shows the example of the metadata in the case of the Tagalog questionnaire document.

Language name:	Tagalog
ISO 639-3:	tgl
Glottolog:	taga12801
Consultant demographics:	A male speaker of Isabela Tagalog. He also speaks Ilocano.
Elicitation dates:	20 Dec 2022, 12 Jan 2023

Figure 1: The metadata example for Tagalog

**Translations** The major part of a questionnaire document consists of glossed examples of the sentences that have been translated from the target sentences in the questionnaire given in the contact language. In the examples, the expressions identified as contributing to the relevant modality are boldfaced and glossed as MOD. If more than one translation is available for a target sentence, where possible, further information is provided regarding factors that distinguish them. This includes how the sentences are different in terms of the consultant’s subjective preference, their syntactic properties and effects of additional contextual manipulations. Negative data have also been recorded. When the consultant judged an alternative sentence suggested by the fieldworker to be incompatible with the given context, the sentence is marked with #. When the sentence is ungrammatical regardless of the context, it is marked with \*. Figure 2 shows a portion of the questionnaire document for Akan including translations for target sentences involving deontic possibility.

**Table** The table summarises the findings with respect to the inventory of modal expressions in the language, organized with respect to force, flavor and strength. The table also includes information about additional grammatical properties that distinguish different expressions. An example table for Vietnamese is given in Figure 3.

## 4.2 Accessing and contributing to the dataset

The dataset itself and the questionnaire are available open access on OSF at [https://osf.io/zngd6/?view\\_only=45c6e82e25034962a1f75fd67642f6f4](https://osf.io/zngd6/?view_only=45c6e82e25034962a1f75fd67642f6f4) (anonymous OSF project for peer-review). Researchers are invited to contribute to the dataset. To do this, please create the questionnaire document for the target language using the original questionnaire and contact the authors to include it in the OSF project.

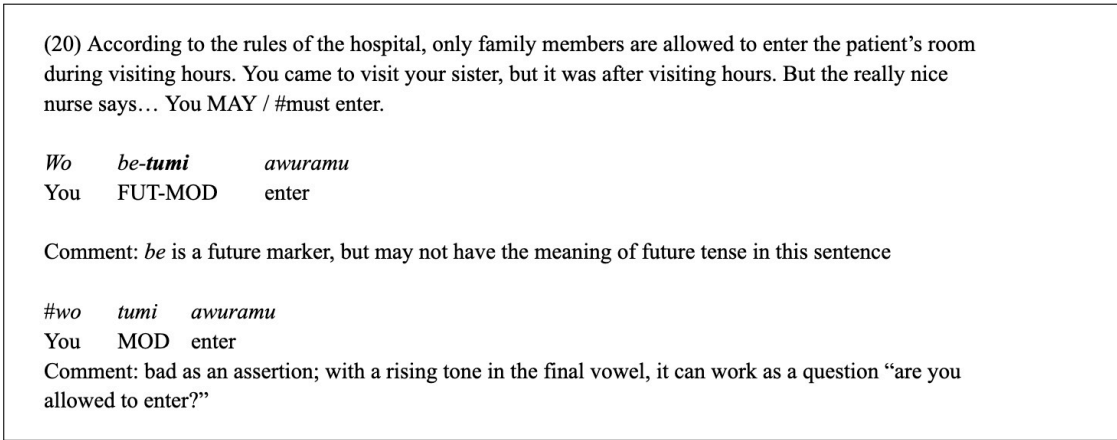


Figure 2: Questionnaire portion for Akan deontic possibility

## 5 Case studies

In this section, we discuss four case studies using our dataset, referring back to the discussion in Sect. 2.2. Specifically, we discuss the following four case studies: (i) assessment of Nauze’s (2008) and Steinert-Threlkeld et al.’s (2023) generalizations regarding force/flavor variability (Sect. 5.1); (ii) explorations of generalizations concerning expressions of non-necessity and impossibility (Sect. 5.2); (iii) classification of the typology of the encoding of modal strength (Sect. 5.3); and (iv) investigation of the role of future in modality (Sect. 5.4).

### 5.1 Generalizations about force/flavor variability

We have investigated the validity of Nauze’s and Steinert-Threlkeld et al.’s IFF generalizations with respect to our data. The generalizations are repeated below:

- (16) **Nauze’s (2008) Generalization**  
 Modal elements can have multiple meanings only along a unique axis out of the force and flavor axes: they either vary with respect to force or they vary with respect to flavor, but they cannot vary on both axes.
- (18) **Independence of force and flavor**  
 All modals in natural language satisfy the independence of force and flavor property: if a modal can express the pairs  $(fo_1, fl_1)$  and  $(fo_2, fl_2)$ , then it can also express  $(fo_1, fl_2)$  and  $(fo_2, fl_1)$ . (Steinert-Threlkeld et al., 2023)

To evaluate these generalizations, we have tested the semantic variability of all modal expressions in the dataset, considering multi-word expressions as one item.

All 24 languages in our sample contained at least one item that exhibits flavor variability. Interestingly, whether the generalizations in (16) and (18) are obeyed in the sample depends on whether strong necessity and weak necessity are considered to be distinct forces for the purpose of the generalizations. If they are, we have a case that is a potential counterexample to the generalizations. Otherwise, the data available in the dataset are compatible with the generalizations.

The pertinent case concerns the Igbo modal *kwesiri*. The item is used in epistemic, deontic

MODALS	EPISTEMIC	ROOT/NON-EPISTEMIC			
		PARTICIPANT-EXTERNAL		PARTICIPANT-INTERNAL	
	Epistemic	Deontic	Teleological	Circumstantial	Bouletic
<b>Necessity</b>	[chắc hẳn (là) // chắc chắn (là)] <sup>SVP</sup> ắt phải (là)/chắc chắn phải (là)	phải <sup>MV</sup> buộc phải	[(cần) phải] <sup>MV</sup>	[cần phải] <sup>MV</sup> // [chắc hẳn] <sup>SVP</sup> // [ắt phải]	Muốn <sup>V</sup> // cần // [cần phải] <sup>MV</sup>
<b>Weak Necessity</b>	[có lẽ (là)] <sup>ADV</sup> chắc là	[nên/?cần phải] <sup>MV</sup>	nên <sup>MV</sup>		...
<b>Possibility</b>	[có thể (là)] <sup>ADV</sup>	[có thể (được)] <sup>ADV</sup> // được (phép) + V // có thể V được	[có thể] <sup>ADV</sup>	[có thể] <sup>ADV</sup> , được <sup>PVD</sup>	
<b>Non necessity</b>	có thể (là) + NEG	NEG + (cần) phải	NEG + cần phải	NEG + chắc chắn	
<b>Impossibility</b>	[chắc chắn (là) + NEG] <sup>SVP</sup> // NEG + thể (nào)	NEG + được (phép)	NEG + thể	NEG + thể	

NEG = không (usually), ch<sup>ả</sup> / ch<sup>ẳ</sup>ng (negation focus)

[MV] Modal Verbs: phải, cần/cần phải, được, nên  
[SVP] Stative Verb/Pseudo-Adjectival Paraphrase: chắc chắn/chắc hẳn/ắt (≠ “sure”) + (là)  
[PVD] Postverbal DUOC: modal ability particle

Figure 3: Summary table for Vietnamese modals

and teleological weak necessity contexts:

- (33) *Context (epistemic weak necessity): You know that Ann goes hiking from 7am until 8am every day. Most of the time Ann goes to the lake, but sometimes she goes to the mountains. It is now 7:30am and your friend asks where Ann is. You say to your friend: Ann should be at the lake right now.*

Ann kwesiri i-no na lake ugbua.  
Ann MOD INF-be.LOC at lake now

- (34) *Context (deontic weak necessity): In England, it is recommended that face coverings be worn in stores, but it is not a legal requirement. You plan on going shopping, and you*

*think to yourself ... I ought to wear a face covering.*

m kwesiri i-yi ihe kpochie ihu m.  
1SG MOD INF-wear thing cover face my

- (35) *Context (teleological weak necessity): There are different ways to get to the shopping centre. You can go by foot, by bus, or drive yourself. You are not sure how to go. Your friend advises you to get the bus because it is cheaper... To get to the shopping centre, you should take the bus.*

i kwesiri i-ji bus gaa.  
2SG MOD INF-take bus go

Furthermore, *kwesiri* can be used to express circumstantial strong necessity:

- (36) *Context (circumstantial strong necessity): You are on a bus. You have not had a chance to go to the toilet for 4 hours, and your bladder is full. You text your friend... I have to pee so badly!*

m kwesiri i-nyu mamiri ozogbo-ozogbo.  
1SG MOD INF-go urine immediately

- (37) *Context (circumstantial strong necessity): Water is entering a pipe, which has two outputs: the left output and the right output. You block the left output so that the water must come out from the right output.*

Mmiri ahu kwesiri i-si aka nri puta.  
water DET MOD INF-follow hand food come.out

However, crucially, *kwesiri* cannot express teleological strong necessity:

- (38) *Context (teleological strong necessity): Team A are playing in a tournament. In order to advance... Team A has to beat Team B.*

#Team A kwesiri i-meri Team B.  
Team A MOD INF-win Team B.

The consultant noted that (38) is not really appropriate in the context, and that it means something like “Team A should have won” (e.g. because they played better). Interestingly, this comment not only indicates that a teleological strong necessity reading is not available for *kwesiri* in (38), it also suggests that the modal construction is compatible with past temporal orientation without any morphologically transparent past shifting, in contrast with the English translation provided by the consultant, which contains perfect aspect.<sup>11</sup>

The Igbo translation provided for the teleological necessity example in (38) is the following, which involves the element *rírí*, to be discussed in Sect. 5.4:

- (39) Team A kwesi-rírí i-meri Team B.  
Team A MOD-MOD INF-win team B

Thus, the distribution of *kwesiri* can be summarized as follows (— indicates that the judgment was not elicited):

- (40) **The distribution of the Igbo *kwesiri***

<sup>11</sup>Thanks to an anonymous reviewer for encouraging us to make this observation explicit.

	Epistemic	Deontic	Teleological	Circumstantial
<b>Strong necessity</b>	—	—	#	✓
<b>Weak necessity</b>	✓	✓	✓	—

If the strong vs. weak necessity distinction is considered to be a relevant distinction for the generalizations in (16) and (18), the Igbo modal *kwesiri* will be a counterexample. It is a counterexample to Nauze’s generalization in (16) because it exhibits both force and flavor variability. It is a counterexample to the IFF in (18) since, even though it expresses circumstantial strong necessity and teleological weak necessity, it does not express teleological strong necessity.<sup>12</sup>

The item will not count as a counterexample if the strength distinction is not relevant for the generalizations. In fact, all modal elements in our data conform to the generalizations if the only relevant force distinction is that of necessity vs. possibility.

## 5.2 Negative modality

In the languages we investigated, we observe that non-necessity is always realized as a combination of morphologically overt negation and a modal marker, while impossibility is lexicalized in several languages in our sample. Moreover, the languages we examined seem to differ in the restrictions on lexicalization induced by modal flavor. Basque and Turkish, for instance, have a lexicalized impossibility modal that occurs with all flavors. More commonly, however, we find languages in which the relevant lexical item is restricted to *deontic* impossibility, for instance in Hausa, Hebrew, and Hungarian. Below we illustrate our findings with examples from Basque and Hausa. Example (41) shows that non-necessity involves separate realization of negation and modality in both languages, again illustrated with deontic flavor.

(41) *Context (deontic non-necessity): In England, it is recommended that face coverings be worn in stores, but it is not a legal requirement. You plan on going shopping, and you think to yourself... I ought to wear a face covering, but I don’t have to wear a face covering.*

- a. **Ez** dut maskara eraman **behar**. (Basque)  
 NEG AUX mask wear MOD(□)
- b. **Ba dole ba** ne in saka kyallen rufe fuska. (Hausa)  
 NEG MOD(□) NEG COP 1SG wear face.covering

In the case of deontic impossibility, by contrast, both Basque and Hausa appear to encode the negative and the modal meaning components in a single lexical item, i.e. *ezin* in Basque and *kada* in Hausa.

(42) *Context (deontic impossibility): You are going to visit your friend in the hospital. When you enter into the hospital, you stop at the information desk to inquire what room your friend is in. But the woman at the information desk tells you that you can’t visit your*

<sup>12</sup>Notably, Steinert-Threlkeld et al. (2023) also consider a weaker version of IFF which is defined in terms of the notion of connectedness. In this version of the IFF, a modal can express the pairs (fo<sub>1</sub>,fl<sub>1</sub>) and (fo<sub>2</sub>,fl<sub>2</sub>) without expressing (fo<sub>1</sub>,fl<sub>2</sub>) as long as it also expresses (fo<sub>2</sub>,fl<sub>1</sub>). In other words, a modal can cover an L-shaped region in the force/flavor table. Whether the Igbo *kwesiri* satisfies this weaker generalization depends on the status of weak circumstantial necessity, which was not investigated in the data collection. If *kwesiri* can express weak circumstantial necessity, it satisfies the weaker IFF, but not the strict IFF defined in (18).

friend now because it's already 8pm. She says: "I'm sorry, the hospital regulations say that... **Visitors mustn't stay after 6pm.**"

- a. Bisitariak **ezin** dira 6 ostean gelditu. (Basque)  
 Visitors MOD( $\neg\Diamond$ ) be.3PL 6 after stay
- b. **Kada** maziyarda su wuce karfe 6 na yamma. (Hausa)  
 MOD( $\neg\Diamond$ ) visitors 3PL stay hour 6 pm

However, the languages differ when it comes to other flavors of impossibility, such as epistemic impossibility as illustrated in (43). In this context, the same impossibility modal *ezin* can be used in Basque (43-a).<sup>13</sup> In Hausa, by contrast, the use of the *kada*-construction in (43-b) is infelicitous (as indicated by the symbol '#'), and impossibility is instead expressed by a transparent combination of negation (*ba ... ba*) and the possibility modal *yiwu*.

(43) *Context (epistemic impossibility): Ben goes swimming every day. Ben is not obliged or required to go swimming; it is just a habit of his. It is now time for Ben to be swimming, so... Ben can't be at home.*

- a. Benat **ezin** da etxean egon. (Basque)  
 Ben MOD( $\neg\Diamond$ ) AUX home be
- b. #**Kada** Ben ya kasance a gida. (Hausa)  
 MOD( $\neg\Diamond$ ) Ben 3SG be at house
- c. **Ba** zai **yiwu** Ben ya kasance a gida **ba**. (Hausa)  
 NEG FUT.3SG MOD( $\Diamond$ ) Ben 3SG be at house NEG

To summarize our empirical results, all languages in our sample behave uniformly regarding non-necessity in that this meaning is not lexicalized. We furthermore observe an asymmetry with respect to impossibility: our data suggest that lexicalization of impossibility meanings may vary with modal flavor. Table 2 summarizes the two patterns of lexicalized impossibility that we identified in our language sample. The first pattern is exhibited by Basque and Turkish, these languages have impossibility modals that are compatible with all modal flavors. The second pattern was found in Hausa, Hebrew, Hungarian, Russian, and Thai, where possibility modals are restricted to deontic flavor.

Modal flavor	'all flavors' impossibility modal	'deontic only' impossibility modal
epistemic	✓	×
teleological	✓	×
deontic	✓	✓
circumstantial	✓	×

Table 2: Lexicalization of impossibility in different flavors: ✓ means the meaning is lexicalized, × means it is not

Thus, the data are consistent with the generalization about the relative complexity of the O-corner modality, repeated below from Sect. 2.2.2.

(44) **Relative complexity of the O-corner modality**

<sup>13</sup>It is likely that the Basque *ezin* is etymologically related to the negative auxiliary *ez*. However, the exactly morphological makeup of *ezin* is non-transparent.

An expression for the O-corner modality is at least as morpho-syntactically complex as modal expressions for the other three corners, i.e., possibility, necessity, and impossibility.

In addition, in our sample there is no language that lexicalizes impossibility but does not lexicalize deontic impossibility. This observation leads us to hypothesize a new generalization in (45) (Uegaki et al., 2023).

- (45) **Deontic Priority generalization:** if a language lexicalizes any impossibilities, then it lexicalizes deontic impossibility.

### 5.3 Encoding of modal strength

All of the four encoding patterns for the strength distinction, discussed in Sect. 2.2.3 and repeated below in (46), are attested in the languages in the dataset.

- (46) A. Lexical distinction  
 B. Strong necessity + additional morphology  
 C. Comparative paraphrase  
 D. Unmarked distinction

Table 3 summarizes the classifications for the 24 sample languages.

Strength encoding type	Languages
A	<i>Akan</i> , Cantonese, Hausa, Hindi, Igbo, <i>Japanese</i> , Khmer, Mandarin, Russian, Thai, <i>Vietnamese</i>
B	Dutch, Greek, Hungarian, Mapudungun, Spanish, <i>Vietnamese</i>
C	Farsi, Hebrew, <i>Japanese</i> , <i>Turkish</i>
D	<i>Akan</i> , Basque, Farsi, Tagalog, Telugu, <i>Turkish</i>
A*	Kiitharaka
A**	Korean

Table 3: Classification of modal strength encoding in the 24 languages. A denotes **lexical distinction**, B **strong necessity + additional morphology**, C **comparative paraphrase** and D **unmarked distinction**, as discussed in Sect. 2.2.3. 4 languages (*Akan*, *Japanese*, *Turkish* and *Vietnamese*, marked in italics) exhibit multiple patterns for different strong/weak necessity pairs. The Kiitharaka pattern A\* and the Korean pattern A\*\* can be considered as yet other types, as discussed in the main text.

Among the observed strength distinctions, two cases of type A (notated with A\* and A\*\* in Table 3) are worth highlighting. In Kiitharaka, epistemic weak necessity is expressed by the item *ũmba*, as in (47):

- (47) *Context (epistemic weak necessity): You know that Ann goes hiking from 7am until 8am every day. Most of the time Ann goes to the lake, but sometimes she goes to the mountains. It is now 7:30am. You say to your friend: **Ann should be at the lake right now.***<sup>14</sup>

<sup>14</sup>Glossing conventions for Kiitharaka: number *n*: agreement marking for the noun class *n*; *nSM* (where *n* is a number): subject marker for subjects with the noun class *n*, INF: infinitive; LOC: locative; FOC: focus marker;

Ann a-**ûmba** kû-îgua a-rî î-ria-ni.  
 Ann 1SM-MOD INF-be 1SM-be 5-lake-LOC  
 “It is likely that Ann is at the lake”

This contrasts with epistemic strong necessity, which is expressed by the *wa + (mw)anka* combination:

- (48) *Context (epistemic strong necessity): Ben goes swimming every day. Ben is not obliged or required to go swimming; it is just a habit of his. It is now time for Ben to be swimming... so Ben must be swimming (right now).*

Ben n-**wa** mw-**anka** wîgue akîbutagîra thaa ino.  
 Ben FOC-MOD 3-MOD be 1SM.swim.PROG.APPL time this  
 “Ben must be swimming right now”

Prima facie, this indicates that we simply have a lexical distinction between strong and weak epistemic necessity. However, the complete picture may not be so simple, as *ûmba* seems to express possibility at least at some level. This is evidenced by the fact that impossibility is expressed by *ti* (negation) + *ûmba* across epistemic, teleological and circumstantial flavors. The case of epistemic impossibility is shown below:

- (49) *Context (epistemic impossibility): Ben goes swimming every day. Ben is not obliged or required to go swimming; it is just a habit of his. It is now time for Ben to be swimming...so Ben can't be at home.*

Ben a-**ti-ûmba** kw-îgua a-rî mû-ciî.  
 Ben 1SM-NEG-MOD INF-be 1SM-be 3-home  
 “Ben can't be at home”

An additional complication is the fact that *ûmba* is not elicited in possibility contexts. Possibility is expressed with the *wa* element (without *mwanka*):

- (50) *Context (epistemic possibility): Charles is looking for his pet cat, but cannot find it. He knows that his cat is very elusive/mysterious. Charles tells his sister: The cat may be inside. The cat may also be outside.*

Kabaka n'**wa** wîgue karî nyomba, kana ûkeegua karî nja.  
 12.cat FOC.MOD be 12.be 9.house or sit.be 12.be 9.outside  
 “The cat may be inside or outside.”

Thus, the data are compatible with a view where an item that is underlyingly a possibility modal (i.e., *ûmba*) undergoes (semantic and/or pragmatic) strengthening to express weak necessity in examples such as (47). We would like to leave further exploration of this possibility for future research.

In Korean, the strength distinction within deontic necessity can be encoded by combining the element *ya* with different particles: *-toy* and *-keyss*. This is exemplified below:

- (51) *Context (deontic strong necessity): In Indonesia, the law states that when you ride a motorbike... You must wear a helmet.*

heylmeys-ul sse-**ya** **toyn**-ta.  
 helmet-ACC wear-MOD MOD.good-DECL

---

PROG: progressive; APPL: applicative.



- (52) *Context (deontic weak necessity): Your friend is having a birthday party. She tells you that she would like it if everyone invited wore formal clothes, but that it is not necessary to wear formal clothes to attend. You think to yourself... I ought to wear formal clothes to the party.*

phathi-ey cengcang-ul ip.e-ya keyss-ta.  
party-DAT formal.clothes-ACC wear-MOD MOD-DECL

Chung (2019) among others argues that Korean deontic necessity and possibility can be decomposed into a conditional element (*ya* ‘only if’ and *to* ‘even if’) and an evaluative predicate (*toy* ‘good’), which is confirmed in our data. It is possible that the strength distinction exemplified above can be accounted for by a decompositional analysis of the same spirit.

#### 5.4 How future contributes to modality

The data from our questionnaires can provide insights beyond the interactions of modal force/strength, modal flavor and negation that are directly tested in the questionnaire. An example of this is the distribution of ‘future markers’ in cross-linguistic modal paradigms. In many languages, different modal contexts elicit morphological forms that are canonically used to express predictive future meanings. Consider, for instance, the Hausa sentence that was elicited in context (43), repeated below.

- (43) *Context (epistemic impossibility): Ben goes swimming every day. Ben is not obliged or required to go swimming; it is just a habit of his. It is now time for Ben to be swimming, so... Ben can't be at home.*

Ba **zai** yiwu Ben ya kasance a gida ba. (Hausa)  
NEG FUT.3SG MOD(◇) Ben 3SG be at house NEG

The translation of the target sentence contains the morpheme *za*, which is most commonly used to express future meaning as in example (53) (volunteered by our consultant), and is described as a future TAM marker in Hausa reference grammars (e.g. Newman 2000).

- (53) **Za** ka saka hular kwano. (Hausa)  
FUT 2SG wear helmet  
“You will wear a helmet.”

It is well known that future markers in many languages are compatible with epistemic modal readings, and in this case typically display (strong or weak) necessity in terms of modal force (see e.g. Giannakidou & Mari 2018; Ippolito & Farkas 2022; Frana & Menéndez-Benito 2023 for recent accounts of epistemic future in Italian). Beyond this, however, the range of modal environments in which future markers occur across languages poses interesting questions concerning cross-linguistic variation in the semantics of (modal) future elements, and the role of so-called ‘future markers’ in the composition of modal meaning more generally. In the Hausa example in (43), for instance, *za* combines with the possibility modal *yiwu*. The exact semantic contribution of *za* is somewhat opaque: it does not seem to contribute modal force, since the sentence obtains a reading of (negated) possibility. However, *za* does not introduce a temporal future shift, either. The target sentence in (43) makes a claim about Ben’s current whereabouts at the utterance time, i.e. it has a present temporal perspective and a present temporal orientation. Note moreover that possibility readings with *za* are not restricted to epistemic flavor or negative

polarity. (54) shows an example in which *za* combines with a different possibility modal (*iya*) in a circumstantial possibility context.

- (54) Context (circumstantial possibility): Ben was in a motorbike accident 3 weeks ago, and he sprained his ankle. Ben is able to walk now. However, the doctor told Ben that he is not allowed to walk until 5 weeks after the accident... **Ben can walk now.**

Ben **zai** iya tafiya yanzu. (Hausa)  
Ben FUT.3SG MOD walk now

Similarly interesting is the distribution of the future marker *ga* in Igbo. Like *za* in Hausa, Igbo *ga* expresses plain predictive future (55). It also patterns with future markers in many Indo-European languages (including English *will*) in that it can express non-future epistemic necessity (56).

- (55) Context question: Where will John be tomorrow at noon?

John **gà-ánò** n'ùlòák wúkwó. (Igbo)  
John FUT-be at.school  
“John will be at school.”

- (56) Context question (epistemic, present oriented): John is not at home, where can he be?

John **gà-ánò** n'ùlòák wúkwó.  
John FUT-be at.school  
“John will be at school.”

Examples such as (56) show that *ga* can express modal meaning without an obligatory future interpretation. However, the modal use of *ga* is not restricted to epistemic flavor or necessity contexts. Our elicitations revealed that *ga* is compatible with non-epistemic necessity (e.g. deontic necessity as in (57)) when it combines with the morpheme *riri*.

- (57) Context (deontic necessity): In Indonesia, the law states that when you ride a motorbike ... **You must wear a helmet.**

Í **gà-éyì** rírí helmet.  
2SG FUT-wear RIRI helmet

Moreover, *ga* occurs in possibility contexts, in combination with the morpheme *ni*.

- (58) Context (circumstantial possibility): Ben was in a motorbike accident and he sprained his ankle. Ben is able to walk now. However, the doctor told Ben that he is not allowed to walk until 5 weeks after the accident... **Ben can walk now.**

Ben **gà-àgá-ní** íjè ùgbúà.  
Ben FUT-go-NI walk now

There is a notable similarity between the Igbo example in (58) and the Hausa example in (54). In both cases, a future modal element combines with a morpheme that seems to restrict the modal force of the overall sentence to possibility (i.e. *iya* in Hausa and *ni* in Igbo). Whatever the exact semantic contribution of the future markers in these cases,<sup>15</sup> a viable and potentially

<sup>15</sup>For the case of Igbo in particular, this question is addressed in more detail in Mucha et al. (to appear), including a sketch of a formal analysis.

interesting hypothesis is that the role of the future modal in possibility contexts is similar in both languages.

As these examples illustrate, our database enables comparison of the distribution of ‘modality-adjacent’ grammatical markers such as future morphemes across systematically constructed modal contexts, potentially leading to the discovery of meaningful cross-linguistic patterns.

## 6 Conclusions

In this paper, we have presented our cross-linguistic dataset of force-flavor combinations in modal elements. We discussed theoretical motivations for constructing the dataset, the data collection methodology, as well as the design and the format of the dataset. Our case studies in four domains illustrate that the dataset supports in-depth assessment of potential cross-linguistic generalizations as well as theory-informed investigations of cross-linguistic variations in modal semantics.

Needless to say, our dataset as it currently stands comes with several crucial limitations. First of all, our language sample is small and is over-represented by specific language families and areas. Furthermore, the methodology of data collection makes it difficult to investigate certain phenomena that fall outside the theoretical framework underlying the questionnaire. For example, investigating graded modality (Lassiter, 2017) is difficult in view of the discrete categorization of force in our data collection methodology. This said, we believe our dataset provides an important first step towards systematic and theory-informed investigation of cross-linguistic generalizations and variation in the domain of modal semantics. It is our hope to collaborate with other scholars to extend the dataset to cover a more diverse range of languages.

## References

- Bar-Lev, Moshe E & Roni Katzir. 2023. Communicative stability and the typology of logical operators. *Linguistic Inquiry* 1–23.
- Bochnak, M Ryan. 2015. Variable force modality in Washo. In *Proceedings of North-East Linguistics Society*, vol. 45, 105–114.
- Cable, Seth. 2017. The expression of modality in Tlingit: A paucity of grammatical devices. *International Journal of American Linguistics* 83(4). 619–678.
- Chung, WooJin. 2019. Decomposing deontic modality in Korean. *Journal of Semantics* 36(4). 665–700.
- Condoravdi, Cleo. 2002. Temporal interpretation of modals: Modals for the present and for the past. In David Beaver, Stefan Kaufmann, Brady Clark & Luis Casillas (eds.), *The construction of meaning*, 59–88. Stanford: CSLI Publications.
- Deal, Amy Rose. 2011. Modals without scales. *Language* 87(3). 559–585.
- Deal, Amy Rose. 2015. Reasoning about equivalence in semantic fieldwork. In M. Ryan Bochnak & Lisa Matthewson (eds.), *Methodologies in Semantic Fieldwork*, 157–174. Oxford University Press.

- von Fintel, Kai & Sabine Iatridou. 2008. How to say ought in foreign: The composition of weak necessity modals. In Jacqueline Guéron & Jacqueline Lecarme (eds.), *Time and Modality*, 115–141. Dordrecht: Springer Netherlands. doi:10.1007/978-1-4020-8354-9\_6. [https://doi.org/10.1007/978-1-4020-8354-9\\_6](https://doi.org/10.1007/978-1-4020-8354-9_6).
- von Fintel, Kai & Sabine Iatridou. 2023. Prolegomena to a theory of X-marking. *Linguistics and Philosophy* 46(6). 1467–1510.
- Frana, Ilaria & Paula Menéndez-Benito. 2023. The evidential future in Italian. *Natural Language Semantics* 31(2–3). 139–178.
- Giannakidou, Anastasia & Alda Mari. 2018. A unified analysis of the future as epistemic modality: The view from Greek and Italian. *Natural Language & Linguistic Theory* 36(1). 85–129.
- Gluckman, John & Margit Bowler. 2020. The expression of modality in Logoori. *Journal of African Languages and Linguistics* 41(2). 195–238. doi:doi:10.1515/jall-2020-2010. <https://doi.org/10.1515/jall-2020-2010>.
- Guo, Qingxia, Nathaniel Imel & Shane Steinert-Threlkeld. 2022. A Database for Modal Semantic Typology. In *Proceedings of the 4th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, 42–51. Seattle, Washington: Association for Computational Linguistics. doi:10.18653/v1/2022.sigtyp-1.6.
- Hacquard, Valentine. 2011. Modality. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An International Handbook of Natural Language Meaning*, 1484–1515. Berlin: Mouton de Gruyter.
- Horn, Laurence R. 1989. *A Natural History of Negation*. Chicago: CSLI Publications.
- Horn, Laurence Robert. 1972. *On the semantic properties of logical operators in English*: University of California, Los Angeles PhD dissertation.
- Imel, Nathaniel, Qingxia Guo & Shane Steinert-Threlkeld. 2023. An efficient communication analysis of modal typology. Unpublished manuscript, University of California, Irvine and University of Washington.
- Ippolito, Michela & Donka F. Farkas. 2022. Assessing alternatives: the case of the presumptive future in Italian. *Linguistics and Philosophy* 45(4). 943–984.
- Jeretič, Paloma. 2021. *Neg-Raising Modals and Scaleless Implicatures*: New York University PhD dissertation.
- Katzir, Roni & Raj Singh. 2013. Constraints on the lexicalization of logical operators. *Linguistics and Philosophy* 36. 1–29.
- Kratzer, Angelika. 1977. What “must” and “can” must and can mean. *Linguistics and Philosophy* 1. 337–355.
- Kratzer, Angelika. 1981. The notional category of modality. In H. J. Eikmeyer & H. Rieser (eds.), *Words, Worlds and Contexts*, Berlin, New York: De Gruyter.

- Kratzer, Angelika. 1991. Modality. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, 639–650. Berlin: De Gruyter.
- Kuhn, Jeremy & Lena Pasalskaya. 2022. Multiple pressures to explain the ‘not all’ gap. Ms., Institut Jean Nicod, ENS, CNRS, EHESS, PSL University.
- Lassiter, Daniel. 2017. *Graded modality: Qualitative and quantitative perspectives*. Oxford University Press.
- Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International journal of American linguistics* 70(4). 369–415.
- Matthewson, Lisa. 2010. Cross-linguistic variation in modality systems: The role of mood. *Semantics and Pragmatics* 3(9). 1–74. doi:10.3765/sp.3.9.
- Matthewson, Lisa. 2011. Methods in cross-linguistic semantics. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, 268–284. Berlin: de Gruyter.
- Matthewson, Lisa. 2013. Gitksan modals. *International Journal of American Linguistics* 79(3). 349–394.
- Matthewson, Lisa. 2016. Modality. In Maria Aloni & Paul Dekker (eds.), *The Cambridge Handbook of Formal Semantics*, 525–559. Cambridge: Cambridge University Press.
- Mucha, Anne, Mary Amaechi, Fred Whibley & Wataru Uegaki. to appear. Future and the composition of modal meaning: the view from Igbo. *Proceedings of Sinn und Bedeutung 28*.
- Nauze, Fabrice. 2008. *Modality in typological perspective*. Amsterdam: University of Amsterdam PhD dissertation.
- Newman, Paul. 2000. *The Hausa Language*. New Haven: Yale University Press.
- Nuyts, Jan. 2006. Modality: Overview and linguistic issues. *The expression of modality* 126. 1–26.
- Peterson, Tyler Roy Gösta. 2010. *Epistemic modality and evidentiality in Gitksan at the semantics-pragmatics interface*: University of British Columbia PhD dissertation.
- Portner, Paul. 2009. *Modality*. Oxford: Oxford University Press.
- Rubinstein, Aynat. 2012. *Roots of Modality*: University of Massachusetts, Amherst PhD dissertation.
- Rubinstein, Aynat. 2014. On necessity and comparison. *Pacific Philosophical Quarterly* 95. 512–554.
- Rubinstein, Aynat. 2021. Weak necessity. In Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann, Thomas Ede Zimmermann & Dina Voloshina (eds.), *The Wiley-Blackwell Companion to Semantics*, 107–132. Oxford: Wiley.
- Rullmann, Hotze & Lisa Matthewson. 2018. Towards a theory of modal-temporal interaction. *Language* 94(2). 281–331.

- Rullmann, Hotze, Lisa Matthewson & Henry Davis. 2008. Modals as distributive indefinites. *Natural Language Semantics* 16(4). 317–357.
- Sardinha, Katie. 2022. The perils of combining translation tasks and judgment tasks. *Semantic Fieldwork Methods* 4(1). Article 6.
- Silk, Alex. 2022. Weak and strong necessity modals: On linguistic means of expressing "a primitive concept ought". In Billy Dunaway & David Plunkett (eds.), *Meaning, Decision, and Norms: Themes from the Work of Allan Gibbard*, 203–245.
- Steinert-Threlkeld, Shane, Nathaniel Imel & Qingxia Guo. 2023. A semantic universal for modality. *Semantics and Pragmatics* 16(1).
- Uegaki, Wataru. 2022. The informativeness/complexity trade-off in the domain of boolean connectives. *Linguistic Inquiry* 1–23.
- Uegaki, Wataru, Anne Mucha, Nathaniel Imel & Shane Steinert-Threlkeld. 2023. Deontic priority in the lexicalization of impossibility modals .
- Vander Klok, Jozina. 2013. Restrictions on semantic variation: a case study on modal system types. In *Workshop on Semantic Variation, University of Chicago*, 25–26.
- Vander Klok, Jozina. 2021. Revised modal questionnaire for cross-linguistic use. Open access field linguistics tool. [http://tulquest.humanum.fr/sites/default/files/questionnaires/70/Modal\\_Questionnaire\\_CrossLing\\_JVK.pdf](http://tulquest.humanum.fr/sites/default/files/questionnaires/70/Modal_Questionnaire_CrossLing_JVK.pdf).
- Vander Klok, Jozina & Vera Hohaus. 2020. Weak necessity without weak possibility: The composition of modal strength distinctions in Javanese. *Semantics and Pragmatics* 13(12). 1–55.
- Weingartz, Siena & Vera Hohaus. to appear. Variable modal strength in Afrikaans and Samoan: Deriving strong necessity from weak necessity. *Proceedings of TripleA 10*.