Investigating the properties of clause-embedding predicates in Polish[*]

Tomasz Klochowicz

University of Amsterdam
klochowicztomasz@gmail.com

Abstract

This paper's main focus lies on the compositional semantics of clause-embedding predicates, i.e. verbs or verb-like expressions that represent a relationship between a subject and a proposition. It provides an improved characterisation of the selectional behaviour of the classes of responsive and anti-rogative predicates in terms of their semantic properties. I propose refinements of the hypotheses by Uegaki and Sudo [US19] and Roelofsen and Uegaki [RU21], that are falsified by empirical data. My new proposals are that *all non-veridical and positively preferential predicates are anti-rogative* and that *all responsive predicates are either Q-to-P or P-to-Q distributive*. The latter hypothesis can be alternatively formulated as: "sentences with corresponding interrogative and declarative complements embedded under the same predicate are always *related by entailment*".

1 Introduction

Our conversations frequently involve the use of clause-embedding predicates, i.e. verbs or verb-like expressions that represent a relation between a subject and a proposition. The compositional semantics of the clause-embedding has been intensively discussed in the recent literature in linguistics and formal semantics [e.g. SE15, TRA18, TRA$^+$19, Ueg19, RU21]. The main objective of studying these predicates is to identify the constraints that describe their selectional behaviour in terms of their semantic properties, i.e to explain why *responsive* predicates like "*to know*" can embed all types of clauses as in (1), but *rogative* predicates, like "*to wonder*" in (2) or *anti-rogative* like "*to believe*" in (3) can embed respectively only interrogative and only declarative clauses. Since clause-embedding predicates occur in many languages, the postulated constraints should be verified against cross-linguistic data.

(1) Ann **knows**...
   a. that Bob left.
   b. whether Bob left.
   c. who left.

(2) Ann **wonders**...
   a. *that Bob left.
   b. whether Bob left.
   c. who left.

(3) Ann **believes**...
   a. that Bob left.
   b. *whether Bob left.
   c. *who left.

As a contribution to this line of research, using the methodology by [URR$^+$22], I created a database of clause-embedding predicates in Polish containing more than 1500 data points and analysing 48 predicates with respect to 23 semantic and 13 selectional properties.[1] Moreover, I verified the constraints postulated in the literature against it. In the remainder of this paper, I will discuss the data collection method and the results of that procedure.

---

[1]The database is available at https://docs.google.com/spreadsheets/d/1qijQdak7ppgDqFB91iY-003P3MV-21Wq/edit?usp=sharing&ouid=117127311628580729011&rtpof=true&sd=true and is a part of my master thesis project [Klo22].

2   Some remarks on the data collection process

To collect the data I used the MECORE questionnaire by Uegaki et al. [URR$^+$22]. It consists of a list of clause-embedding predicates in English and several linguistic tests which ought to be performed on each predicate to analyse its semantic and selectional properties. Moreover, the questionnaire provides a spreadsheet to summarise the collected data. Therefore data collection process can be described in three steps: 1. translation, 2. application of the tests, and 3. testing and adjusting the tests and predicates.

The database, which is the main output of this process is a table with properties as columns and predicates as rows, which allows to easily check the validity of a hypothesis describing the behaviour of the clause-embedding predicates. To evaluate a hypothesis in the form of an implication, we isolate all the verbs which satisfy the property from the antecedent and check whether they also satisfy the property from the consequent. To check the validity of a sentence in the form of a bi-implication, we do this twice, once for each implication. To assess the validity of a universal statement, we check whether all the verbs satisfy the relevant property. These procedures will automatically result in finding the relevant counterexamples if they exist. This method can be very fruitful and provide interesting results, which I discuss the next section. However, before proceeding to this part, I need to address its several shortcomings and issues.

The length   of the list of predicates makes it infeasible to run a study with enough native speakers to ensure full-proof results. Therefore, the results of the study are questionable as they depend on the judgment of one person. Since the database contains around 1500 data points mistakes are statistically unavoidable. Moreover, the judgments frequently differ between native speakers, and this method will not detect any variation. To accommodate this issue, the primary investigators organised regular meetings where consultants working on data from different languages could compare their results. Obviously, cross-linguistic variations exist, but since the predicates have the same meaning, they should not be widespread. Therefore, looking at some language can provide insight into another, which helps avoid some mistakes.

The confirmation   of the hypotheses, both existing in the literature and new ones is questionable. Since the database does not include all predicates and since only one consultant creates it, it is impossible to get an ultimate confirmation of a hypothesis even for one language. Moreover, because of these issues the actual degree of confirmation is rather low. However, the study can provide a cross-linguistic confirmation of the counterexamples of some hypotheses and strengthen their falsification or even allow to find new counterexamples as presented in the next section. Moreover, running a short study on a bigger sample can be quickly performed to confirm or reject the falsifying piece of evidence as it needs to consider only one predicate and a few relevant properties.

The generality   of the claims in the questionnaire is questionable. By definition, any semantic property ranges over all possible subjects and all possible complements. The consultants of the questionnaire analyse only one specific example, which may lead to false conclusions for some verbs. Therefore a strong (but fairly standard) assumption is required, namely that the predicates behave uniformly concerning all the sufficiently similar subjects and complements [Klo22].

The scope   of the study includes only some semantic and selectional properties. The choice of those properties is inspired by previous research. Therefore discovering a new hypothesis

in the process is unlikely, as the combinations of the properties have already been checked for English and other languages. Considering the most influential papers (e.g. [SE15], [TRA$^+$19] or [RU21]), the most fruitful way to describe the behaviour of clause-embedding predicates is to look at the properties that haven't been considered before. Moreover, the study only analyzes the selectional properties of the verbs in simple contexts like $\ulcorner sVc \urcorner$ where s is a subject, V is a predicate and c is a complement of some type. Many verbs change their selectional properties in more complex contexts (under negation, modalities or tense/aspect change). For instance, the English predicate "*to doubt*" usually does not take interrogative complements, but in some negative contexts, it does as in (4). This behaviour should also be explained by the semantics theory of clause-embedding predicates.

(4)   "Nobody doubts who the mayor of London is; nobody doubts who speaks for Scotland...".[2]

It should be clear by now that all the issues are interdependent. For instance, performing a study that would provide a higher degree of confirmation would drastically increase the size. Therefore the only reasonable approach is to keep all these three factors balanced, and such a balance seems to be achieved by the current method. Moreover, considering additional predicates, properties, or even contexts can be done without trouble, but it is impossible to include everything.

Is it, therefore, worth the effort to perform such a study? I would argue that it is. First of all, as mentioned above, the database is a good falsification tool. It allows detecting potential counterexamples and running a detailed study of these predicates to confirm the consultant's judgement. Thus the database should not be treated as an oracle that gives the researchers written-in-stone facts about language. However, it can provide a good starting point for further independent research. Therefore the size of the database can be seen as its feature because it allows looking at many predicates to determine which are worth to be investigated closely. The next section provides an example of the important conclusions, which can be drawn from the database.

3   Modelling selectional behaviour

In this section, I will discuss the results of the empirical study of clause-embedding predicates in Polish. I will present some insights into the veridicality hypothesis proposed by Paul Egré [Egr08], and afterwards, I will present empirical evidence which falsifies two other hypotheses previously postulated in the literature and propose their refinements.

3.1   The veridicality hypothesis

A predicate (V) is called *veridical* with respect to declaratives if and only if a sentence which relates a subject s to a declarative complement d using V implies the complement d. For instance, the English predicate "to know" is veridical with respect to declaratives as (5-a) implies (5-b):

(5)   a.   Ann knows that Peter left.

---

[2]https://www.theguardian.com/uk-news/2016/feb/11/labour-tony-lloyd-greater-manchester-mayoral-bid-devolution [accesed: 13.05.2022] Found using the *English corpus of News on the Web* [Dav16]. See also other examples for the predicates "*believe*", "*be certain*", "*think*", "*hope*" in the articles: [Rob19], [May19], [Özy21] and [Whi21].

       b.   Peter left.

The veridicality hypothesis $\mathbb{V}$ proposed by Paul Egré claims that all veridical predicates are responsive [Egr08]. Egré himself points out that emotive predicates like "*to regret*" are potential counterexamples to his hypothesis, as they seem to be veridical and not embed questions. Mayr adds the verb "*to resent*" to this list [May19].

However, Egré argues that these emotive predicates should not be seen as a counterexample to his hypothesis, as they only imply that the complement is believed by the subject and not that it is true [Egr08]. However, the evidence suggests that they pass the veridicality test so they can be seen as proper counterexamples. I will argue that further empirical research should be performed to check the validity of this hypothesis.

Intuitively veridical predicates should not allow for complement cancellation, as they imply that the complement is true. Therefore, the empirical study, except just checking the assertability of sentence (7-a), should also test whether sentences like (7-b) or (7-c) are well-formed. In these cases the intuitions of the native speakers are clear, and most of them claim that there is something wrong with these sentences, as they seem to be contradictory. However, some are still convinced that these sentences are more or less correct. Thus investigating these sentences may lead to interesting conclusions regarding the verb "*to regret*".

(6)   Peter invited Anna on a date. His evil twin - Paul - decided to destroy his brother's relationship and showed up at the meeting with Anna instead of Peter. During the evening, Paul was very rude, and thus Ann regrets that this meeting happened. However, Ann is still convinced that she met with Peter, not Paul.

(7)   a. Ann regrets going on a date with Peter.
         b. ?Ann regrets going on a date with Peter, but in fact, she went on a date with Paul.
         c. ??Ann went on a date with Paul, but she regrets going on a date with Peter.

Investigation of predicates like "*regret*" may also lead to interesting conclusions regarding the property of veridicality. As mentioned above, Egré's motivation to discuss this property is the fact that it is the main semantic difference between "*know*", which is responsive and "*believe*" or "*think*", which are anti-rogative in neutral contexts [Egr08]. It seems that "*regret*" lies somewhere on the spectrum, which starts with a non-veridical "*think*" and ends with the veridical "*know*". Perhaps it is at one extremum or the other, but the evidence suggests that it is somewhere between the two ends. The future study should consider it possible that "*regret*" may undermine the assumptions of the MECORE project, i.e. that all the semantic properties are binary.

3.2   Analysis and refinement of the $\mathbb{NVP}$ hypothesis

The $\mathbb{NVP}$ hypothesis proposed by Ueagki and Sudo claims that any predicate that is non-veridical and preferential is also anti-rogative [US19]. There are at least three predicates in Polish that are counterexamples to this hypothesis: "*martwić się*", "*bać się*" and "*narzekać*" (*eng*: "*be worried*", "*fear*" and "*complain*" respectively).

To derive the anti-rogativity of the non-veridical preferentials Ueagki and Sudo proposed that an utterance of such a predicate presupposes *Threshold Significance*, i.e. that "[there is] an element in the comparison class whose degree along the relevant scale [of preference] is higher than the threshold [of preference]"[US19, p.335]. Observe that, all the counterexamples violate this presupposition, as they are *negatively* preferential. For instance, in the context (8), the premise of the reasoning (9) is true, but the conclusion is not. Thus, since the predicate

to be worried can be meaningfully uttered in this context, it does not presuppose *Threshold Significance*.

(8)   Peter is a prince who needs to marry the winner of the race. He knows nothing about the participants but it still matters to him who will win the race.

(9)   Peter is worried who won the race.
        $\not\leadsto$ Peter is worried that x_F won the race.

To accommodate for the issues with the $\mathbb{NVP}$ hypothesis, which are caused by predicates "martwić się", "bać się" and "narzekać", a different conclusion should be drawn from the derrivation by Uegaki and Sudo [US19]. Observe that the *Threshold Significance* presupposition coincides with the property of being *positively preferential*, which is that, in the declarative case, an agent prefers the clause embedded under the predicate over the other options. Therefore I argue that the condition of being *positive* should be added to the $\mathbb{NVP}$ hypothesis. The new claim is not violated by my counterexamples and ensures that the predicates that satisfy the antecedent of the hypothesis also satisfy the *Threshold Significance* presupposition. Thus I propose the hypothesis $\mathbb{NVPP}$ as a refinement of the hypothesis $\mathbb{NVP}$. $\mathbb{NVPP}$ is formally justified by the derivation in the aforementioned paper and independently confirmed by the data collected in my database:

Hypothesis 1. $\mathbb{NVPP}$: *All non-veridical and positively preferential predicates are anti-rogative.*[3]

3.3   The entailment hypothesis

The $\mathbb{PQ}$ hypothesis proposed by Roelofsen and Uegaki claims that all responsive predicates are P-to-Q distributive, i.e. that a sentence with a declarative complement embedded under a responsive predicate entails the same sentence but with a respective interrogative [RU21]. For example, the English predicate "to announce" is P-to-Q distributive as (10-a) implies (10-b). Q-to-P distributivity can be defined as an inverse of this implication.

(10)   a.   About some player a Alfred announced that a won the race.
        b.   Alfred announced who won the race.

The analysis of the database shows that there are verbs that are not P-to-Q distributive. Observe that the verb "*to learn*" seems to behave the same way in English. For instance, the sentence (11), which has a false declarative complement, can be meaningfully uttered, but (12) seems to presuppose that Ann learned the *true* answer to the question. Thus (11) is inconsistent with (12). Thus the predicate "*to learn*" violates the $\mathbb{PQ}$ hypothesis.

(11)   At school, Ann learned that Benjamin Franklin was the first president of the USA.

(12)   At school, Ann learned who was the first president of the USA.

The issue is even more visible in Polish. The context (13) clearly shows that a subject can *learn* false information and that even thought, the sentence (14) contains a question, which the sentence (13) answers, it is not assertible in this context.

---

[3]A similar hypothesis was also independently suggested in the GLOW presentation by Özyıldız et al. as a potential solution to the issue with negative preferential predicates [ÖQR$^+$22].

(13) a. "Constantin Reliu dowiedział się w styczniu, że ... nie żyje. Próbuje co prawda udowodnić rumuńskim władzom, że jest inaczej, ale napotyka potężny opór biurokracji".[4]

      b. In January, Constantin Reliu learned that ... he is dead. He tries to prove to the Romanian administration that it is not the case, but he is met with strong resistance on their side.

(14) Constantin dowiedział się   czy   nie   żyje.
      Constantin learned   himself whether NEG he is alive.
      "Constantin learned whether he is dead."

However, it is important to notice that in the database there is no predicate that would be neither Q-to-P nor P-to-Q distributive. Moreover, Spector and Egré independently claimed that the existence of predicates, that would have a completely unrelated meaning when they embed declaratives and interrogatives is implausible [SE15].[5] From this observation, they conclude that it is always possible to define the meaning of any sentence containing some embedded interrogative by the meaning of the respective declarative embeddings. I would argue that this observation cannot be strengthened like this and, based on the database, I would like to propose the following hypothesis:

Hypothesis 2. *(Entailment hypothesis) All responsive predicates are Q-to-P or P-to-Q distributive.*

This hypothesis can be alternatively formulated as: "sentences with corresponding interrogative and declarative complements embedded under the same predicate are always *related by entailment*" and as such it follows the idea of [SE15]. Moreover, is also a weakening of the $\mathbb{PQ}$ hypothesis and accommodates all the aforementioned counterexamples to it. In particular, it predicts the existence of predicates that are non-veridical but Q-to-P veridical, which are considered by Roelofsen and Uegaki to be potential counterexamples to their hypothesis (Buryat "*hanaxa*" and Turkish "*bil*") [RU21].[6]

4   Conclusion

In this paper, I presented a database of the clause-embedding predicates in Polish and the main conclusions arising from its analysis. I provided new linguistic evidence regarding the predicate "*to doubt*", discussed the non-standard behaviour of the predicate "*to regret*", and proposed two refinements of the hypotheses that were previously established in the literature. I argued that *all non-veridical and positively preferential predicates are anti-rogative* and that *all responsive predicates are either Q-to-P or P-to-Q distributive*, which can also be formulated as "sentences with corresponding interrogative and declarative complements embedded under the same predicate are always *related by entailment*." These hypotheses provide a better understanding of the behaviour of clause-embedding predicates and are a step towards their full unified theoretical analysis.

---

[4]https://www.rmf24.pl/fakty/swiat/news-w-styczniu-dowiedzial-sie-ze-nie-zyje-wladze-nie-chca-przyja,nId,2558031#crp_state=1 [access: 01.06.2022]

[5]e.g. their "*shknow*", which they define as "know" if it embeds a declarative and as "wonder" if an interrogative.

[6]However, as pointed out by [RU21] the predicate "*magtaka*" in the Tagalog language is translated to English as "to surprise" with declarative complements and as "wonder" while embedding interrogatives. A more detailed study of this predicate is needed to investigate whether it indeed violates $\mathbb{E}$.

References

[Dav16]   Mark Davies. Corpus of news on the web (NOW). Available online at https://www.english-corpora.org/now, 2016.

[Egr08]   Paul Egré. Question-embedding and factivity. *Grazer Philosophische Studien*, 77(1):85–125, 2008.

[Klo22]   Tomasz Klochowicz. Investigating semantic and selectional properties of clause-embedding predicates in Polish. Available at: https://www.illc.uva.nl/Research/Publications/Reports/MoL/ in the *Master of Logic (MoL) Series*, 2022. Institute for Logic, Language and Computation, Universiteit van Amsterdam.

[May19]   Clemens Mayr. Triviality and interrogative embedding: context sensitivity, factivity, and neg-raising. *Natural Language Semantics*, 27(3):227–278, 2019.

[ÖQR⁺22] Deniz Özyıldız, Ciyang Qing, Floris Roelofsen, Maribel Romero, and Wataru Uegaki. Cross-linguistic patterns in the selectional restrictions of preferential predicates. Presentation on a work in progress, 2022.

[Özy21]   Deniz Özyıldız. *The event structure of attitudes*. PhD thesis, University of Massachusetts Amherst, 2021. Doctoral Dissertations. 2209.

[Rob19]   Tom Roberts. I can't believe it's not lexical: Deriving distributed veridicality. In *Semantics and Linguistic Theory*, volume 29, pages 665–685, 2019.

[RU21]   Floris Roelofsen and Wataru Uegaki. Searching for a universal constraint on the possible denotations of clause-embedding predicates. In *Semantics and Linguistic Theory*, volume 30, pages 542–561, 2021.

[SE15]   Benjamin Spector and Paul Egré. A uniform semantics for embedded interrogatives: An answer, not necessarily the answer. *Synthese*, 192(6):1729–1784, 2015.

[TRA18]   Nadine Theiler, Floris Roelofsen, and Maria Aloni. A uniform semantics for declarative and interrogative complements. *Journal of Semantics*, 35(3):409–466, 2018.

[TRA⁺19] Nadine Theiler, Floris Roelofsen, Maria Aloni, et al. Picky predicates: Why believe doesn't like interrogative complements, and other puzzles. *Natural Language Semantics*, 27(2):95–134, 2019.

[Ueg19]   Wataru Uegaki. The semantics of question-embedding predicates. *Language and Linguistics Compass*, 13(1), 2019.

[URR⁺22] Wataru Uegaki, Maribel Romero, Floris Roelofsen, et al. Mecore: Cross-linguistic investigation of meaning-driven combinatorial restrictions in clausal embedding, 2022. https://wuegaki.ppls.ed.ac.uk/mecore/ [Accessed: 21.02.2022].

[US19]   Wataru Uegaki and Yasutada Sudo. The *hope-wh puzzle. *Natural Language Semantics*, 27(4):323–356, 2019.

[Whi21]   Aaron Steven White. On believing and hoping whether. *Semantics and Pragmatics*, 14:6, 2021.