

The Semantically Annotated Corpus of Polish Quantificational Expressions

Jakub Szymanik

Institute for Logic, Language and Computation
University of Amsterdam
jakub.szymanik@gmail.com

Witold Kieraś

Institute of Computer Science
Polish Academy of Sciences
Witold.Kieras@ipipan.waw.pl

Abstract: The paper presents a manually annotated corpus of Polish quantificational expressions. The quantifier annotation was conducted on top of existing gold-standard data for Polish as its separate layer. This paper releases the data and gives an overview of the corpus and related tools. As far as we know, this is the first large-scale annotation of generalized quantifiers together with their crucial semantic properties, including monotonicity profile. We also discuss the potential further use of the corpus in linguistics and cognitive science.

Keywords: generalized quantifiers, corpus, annotation

1. Overview

The paper presents a manually annotated corpus of quantificational expressions of Polish. The corpus is a new separate layer of annotation in the gold-standard 1.2 million tokens large subcorpus of the National Corpus of Polish (NKJP1M, Przepiórkowski et al., 2012). It is a balanced set of short samples (approx. 40-60 words long) representing different text genres and available on GNU GPL. It is the most widely used resource for Polish in standard NLP and machine learning tasks covering automatic annotation of various levels. Its annotation features adjudicated sentence- and word-level segmentation, morphosyntactic description, shallow parsing (syntactic words and groups), named entities description, and limited word sense disambiguation. Thus the quantificational layer contributes to the semantic level of the gold-standard annotation of the dataset and, at the same time, may benefit from the existing morphological and syntactic layers.

The paper describes the process of manual annotation of the corpus regarding quantificational expressions and their features, gives an overview of the corpus, and points out its relevance for linguistics and cognitive science. The corpus will also serve as a referential data set and training data for machine learning classifiers.

2. Related work

According to our knowledge, there were no previous large-scale attempts at the manual annotation of generalized quantifiers in any language, so the presented corpus is a pioneering work in the field. However, we are aware of two small-scale quantifier corpus research. Higgins & Sadock (2003) has used the Penn Treebank Annotation of the quantifier phrases to propose a machine learning approach to modeling quantifier scope preferences. Their research resulted in a dataset of 893 double-quantified sentences, annotated with Penn Treebank II parse trees and hand-tagged for the primary scope reading. Reguera & Stender (2013) has conducted a contrastive study of quantifier use in Spanish and German 60 economic texts in online media. As we will see, the corpus being released in this paper has not only a much more comprehensive coverage but is also tagged with very general semantic features of quantifiers.

Two works had a substantial impact on the presented project. The first of the two is an extensive two-volume survey of the quantificational expressions from the cross-linguistic perspective (Keenan and Paperno, 2012; Paperno and Keenan, 2017). In the introductory

chapter *Quantifier Questionnaire*, many useful distinctions, and guidelines for recognizing and describing quantifiers were specified. Polish was not included among 34 languages presented in the survey. The genetically and typologically closest language considered in both volumes was Russian as the only Slavonic language. The other significant source of motivation was a study by Szymanik and Thorne (2017) in which the authors investigate the frequencies of thirty-six most common quantifiers in English in The WaCky corpus (Baroni et al., 2009). The authors have shown that semantic complexity (Szymanik, 2016) contributes to explaining the differences in frequency distributions. The major limitation of this study is the restriction to the small group of English quantifiers. The corpus described in the current paper will allow, for instance, to refine the results of Szymanik and Thorne (2017) by counting all the quantifiers that occur in a corpus with their semantic features, including semantic complexity, and recognizing the more robust statistical patterns.

3. Quantifiers

Quantifiers are semantic objects. Intuitively, by quantifier or quantificational expression, we understand a natural language expression indicating quantities that are topic neutral, i.e., the truth of the quantifier statement does not depend on the particular individuals to be considered. Typical examples in English are All, not quite all, nearly all, an awful lot, a lot, a comfortable majority, most, many, more than n , less than n , quite a few, quite a lot, several, not a lot, not many, only a few, few, a few, hardly any, one, two, three. Extensionally, a quantifier can be represented as a relation between two sets (properties), $Q(A, B)$. For instance, a quantifier “Some As are B” can be thought of as a relation between predicates A and B, specifically to make the sentence true, the extensions of the two predicates need to overlap. Analogously, the meaning of quantifier “All” can be given in terms of inclusion relation of the denotation of the restrictor (first argument) in the denotation of the scope (second argument).¹

Mathematically speaking, there are other possible types of generalized quantifiers; however, quantifiers taking two properties as their arguments (as defined above) are the most common across natural languages (Peters and Westerståhl, 2006). They are also most intensively discussed in semantic literature. There is no agreement among linguists whether,

¹ Formally, a generalized quantifier Q of type $(1, 1)$ is a functional relation associating with each model M a relation between sets in its universe. Additionally, Q needs to be preserved by bijection, meaning that if two models M and M' are isomorphic then they need to satisfy exactly the same quantifier relations (see Peters and Westerståhl, 2006 for a textbook presentation of generalized quantifiers). This last condition guarantees the mentioned above topic neutrality.

and if yes, which of the more complex quantifiers are even expressed in any natural language (Keenan 1992, Beck 2000). Last but not least, by restricting attention to those types of quantifiers, we make the annotators' task practically feasible. We eliminate some commonly occurring expressions that are not prototypical examples of quantifiers but could be interpreted in the framework of generalized quantifier theory. For instance, proper names, like John, are often interpreted as quantifiers of type (1), and possessive are quantifiers not satisfying the topic neutrality condition, i.e., isomorphism (see, e.g., Peters and Westerståhl, 2006 for examples and more extensive discussion).

The annotators' task was to identify a quantifier and describe its three features, which we will define in the following sections. The chosen features were: grammatical structure (D vs. A quantifiers; see Section 3.1), quantificational force (universal, existential, or proportional; see Section 3.2), and its positivity/negativity (monotonicity profile; see Section 3.3). The most important part of the annotation specified each quantifier's features in terms of categories described in the annotators' manual and presented briefly in the following subsections. In selecting the features, we follow Keenan & Paperno's (2012) comprehensive overview of quantifier properties from a cross-linguistic perspective. We have selected those features as they are among the most critical linguistic and logical characteristics of the quantifiers. The first feature informs us about the basic syntactic and predicate properties of the quantifier. The second one roughly characterizes its meaning and complexity. The third one gives information about the inferential and grammatical properties of the quantifiers (e.g., downward monotone quantifiers are known to trigger negative polarity items) (Ladusaw, 1979). These features also play an essential role in the linguistic debate about the characterization of natural language quantifiers and their universal properties (see Section 8 of the paper for some discussion). Furthermore, the annotations should be further extended in the future with other properties of quantifiers. Describing the three selected key dimensions of quantifier meanings will help with such further work.

3.1 D- and A-quantifiers

The first category distinguishes D-quantifiers from A-quantifiers (Bach et al., 1995; Partee, 1995). This feature of quantifiers refers to syntactic and predicate structures in which the quantifier occurs. In the predicate-argument structure of an utterance, D-quantifiers form expressions that are predicates (nominal phrases), but A-quantifiers directly build or modify predicates. This semantic distinction is also reflected in purely syntactic functions of the expressions: D-quantifiers are usually nouns, adjectives, or numerals (ex. (1)), whereas

A-quantifiers are verb modifiers: verbal affixes, auxiliary verbs, or adverbs (ex. (2)). In the context of our corpus, A-quantifiers are almost exclusively adverbial phrases or functionally adverbial idiomatic expressions. Among the most frequent ones are mostly temporal adverbs such as *(nie) zawsze* '(not) always', *nigdy* 'never', *często* 'often', *czasem*, *czasami* 'sometimes', *rzadko* 'rarely'. Also, adverbial phrases indicating repeatability of events are common: *raz* 'once', *dwa razy* or *dwukrotnie* 'twice', *wiele razy* or *wielokrotnie* 'many times'.

(1) Każdy chce być jak najlepszym rodzicem.

Everyone wants to be as best parent.

'Everyone wants to be the best parent possible.'

(2) Dziennikarze wielokrotnie informowali o nieprawidłowościach.

Journalists repeatedly reported about irregularities.

'Journalists have repeatedly reported about the irregularities.'

There exist interpretations of some verbal prefixes *na-* and *po-* as A-quantifiers in Slavonic languages (i.e., Russian, Paperno, 2012), which can also be applied to Polish. In practice, however, they appeared only three times in our corpus even though specific examples of such prefixal quantifiers were explicitly given in the annotation manual, as in the following example of verbal prefix *na-* which has cumulative meaning:

(3) Do pokoju na=wlatywało komarów.

To room na=flew mosquitoes.GEN

'A lot of mosquitoes flew into the room.'

(4) Anna na=piekła ciasta.

Anna na=baked cake.GEN

'Anna baked plenty of cake.'

The reason for the rareness of such constructions in our corpus is that they are rather colloquial, and only a small fraction of our corpus consists of spoken data. In fact, two of the three examples of such prefixal quantifiers appeared in the spoken subcorpus.

3.2 Universal, existential, and proportional

The second category distinguishes between existential (intersective), e.g., *some*, *none* (see examples (5) and (6)), universal (co-intersective), e.g., *all*, (example (7)) and proportional

quantifiers, e.g., *many*, *every third* (examples (8) and (9)). The criteria for distinguishing the three are extensional and adopted after Keenan and Paperno (2017). For Q being a quantifier and A, B sets, if $Q(A, B)$ is determined by $A \cap B$, that is the set of A s that are B s, then Q is existential (intersective). If $Q(A, B)$ depends on the property $A - B$, that is the set of A s that are not B s, then Q is universal (co-intersective). If $Q(A, B)$ depends on the proportion of A s that are B s, that is $|A \cap B|/|A|$, then Q is proportional.

(5) Niektóre wartości uległy w ostatnich latach dewaluacji.

Some values succumbed in recent years devaluation.

'Some values have devalued in recent years.'

(6) W momencie wybuchu pożaru w budynku nie było nikogo z domowników.

At moment outbreak fire in building not was none of members of the house.

At the time of the fire, there were no members of the house in the building.

(7) Wszyscy uczniowie w zasadzie są przeciwko stosowaniu przemocy.

All student generally are against use violence.

'All students are generally against the use of violence.'

(8) Wielu radnych pobiera wysokie diety.

Many councillors take high diets.

'Many councillors are on high diets.'

We also distinguished a class of numeral quantifiers (unmodified numerals), e.g., *5*, which are restricted only to quantifiers expressed by a number. The motivation for the additional value of the category is purely practical and technical: numeral quantifiers are one of the most frequent in texts and relatively less interesting, so marking them with a separate label provides an easy way to filter them out. So far, we did not distinguish among existential quantifiers a separate class of modified numerals (e.g., *more than 5*), which would be a possible future extension. In line with Szymanik and Thorne's (2017) complexity analysis, we expect that existential and universal quantifiers will be the most frequent, followed by the proportional quantifiers.

Among universal quantifiers, a vast majority are D-quantifiers *każdy* 'each' and *wszystko/wszyscy* 'every' followed by an A-quantifier *zawsze* 'always'. The most frequent existential quantifiers are *kilka* 'couple of', *nic* 'nothing', *żaden* 'none', *nikt* 'no one', *jakiś* 'some'. The most frequent A-type existential quantifier is *nigdy* 'never'.

The paradigmatic example of a proportional quantifier is *większość* 'most'. The class also contains quantifiers, like *wiele* 'many' or *mało* 'few'. We are aware that these quantifiers may be sometimes, depending on the context, also interpreted as existential constructions (Partee, 1989). However, as in the majority of cases, the proportional interpretation seems to be available, so we decided to uniformly treat those expressions as proportional and leave the empirical research into other possible meanings for the future. Furthermore, proportional quantifiers often consist of more than one token. A significant number of them are those expressing a percentage of a whole population, such as:

(9) Co trzeci użytkownik mieszka w wielkim mieście, a tylko 9 proc. to mieszkańcy wsi.

Every third user lives in large city, and only 9 percent are inhabitants countryside.

Every third user lives in a large city, and only 9 percent. are inhabitants of the countryside.

We also treated synonymous quantifiers *oba* and *obydwa* 'both' as a special case of proportional quantifiers expressing the meaning: 'two out of two'. They consist of about 10% of all proportional quantifiers in the corpus.

3.3 Monotonicity

The third category described for each quantifier is its left and right monotonicity annotated as two separate features but with the same range of values. Both are tested independently for each quantifier, and the category can take one of three values: increasing, decreasing, and non-monotone. A quantifier Q is upward monotone (increasing) in its left (respectively, right) argument if and only if, for any sets A , B , C , and D , if A is a subset of C and B is a subset of D , then $Q(A, B)$ entails $Q(C, B)$ (respectively, $Q(A, B)$ entails $Q(A, D)$). As the property's value might not be determined directly in the context of a corpus utterance, the annotators were encouraged to use diagnostic sentence schemes for testing the monotonicity of the quantifiers. For example, a quantifier *some* may be put in the following context:

(10) Some students like candy.

(11) Some people like candy.

(12) Some students like sweets.

From the fact that sentence (10) logically entails sentence (11), it can be seen that the quantifier *some* is upward monotone on its left argument. As sentence (10) implies sentence

(12) *some* is also upward monotone in its right argument. Polish quantifier ‘niektóre’ as in sentence (5) displays the same inference pattern. English quantifier, *no*, corresponding, for instance, to Polish ‘nikogo’ from sentence (6) displays the opposite pattern characteristic for left and right downward monotonicity.

(13) No people like sweets.

(14) No students like sweets.

(15) No people like candy.

If a quantifier is not an upward or downward monotone in its left or right argument, e.g., ‘exactly 5’ translated to Polish as ‘dokładnie 5’, then we say that the quantifier is non-monotone in this argument.

Right monotonicity is crucial for semantic and psycholinguistics research. Barwise and Cooper (1981) even proposed it as one of the semantic universals – a property that every language of the world satisfies. The proposed generalization can be formulated as all simple D-quantifiers are the right monotone or are conjunctions of the right monotone quantifiers. The conjunctions of monotone quantifiers are sometimes also called connected quantifiers. Therefore, we expect that all (or almost all) monomorphemic D-quantifiers in our corpus should be right monotone or connected. Furthermore, there is ample psycholinguistic evidence that right downward monotone quantifiers are harder to process for humans (reasoning, comprehension, verification, and acquisition); see, e.g., Szymanik (2016) for an overview or Deschamps et al. (2015) for recent experimental evidence — one possible explanation associates this extra complexity with a lower overall frequency of right downward monotone quantifiers. Our corpus allows directly comparing the frequencies of downward and upward monotone quantifiers.

4. Annotation and tools

Since there were no large-scale semantic attempts at annotating quantifiers so far and no specific guidelines were established, we have decided to follow the general best practices in manual corpus annotation. Each sample in the corpus was annotated simultaneously by two independent annotators. An additional adjudicator resolved conflicts between the two. Since the quantifier theory involves interdisciplinary research originating in logic and linguistics, we have decided to recruit annotators with different backgrounds and divide them into two teams. The first team consisted of cognitive science undergraduate students. Most of them

had no previous experience with linguistic annotation of any kind but had a more substantial logic background. According to the recruitment process, they needed to complete at least four semesters of formal logic courses to be hired in the project. The second team consisted of four qualified linguists (graduates in Polish philology), who were experienced in various linguistic annotations: morphological, syntactic, and semantic, but with no background in logic. Each sample was annotated by one annotator from each team to diversify insights and reduce oversights in the corpus material. One of the authors of this paper (with a background in both linguistics and logic) served as an adjudicator for the whole project, occasionally consulting the other author. The adjudicator resolved many conflicts between the annotators, see the next section for details on the inter-annotator agreement. We believe that by recruiting two teams of annotators with different educational backgrounds, we could better identify all the possible quantificational expressions. For that reason, any future extension of the annotation will be much faster and easier. The annotators also had access to a dedicated mailing list, where they could ask questions and discuss problems concerning their work. Figure 1. presents an example of a collision between two annotators and adjudicator's choice as seen in the WebAnno application.

The annotation was conducted in the web-based application WebAnno (Eckart de Castilho et al., 2016) designed for different types of linguistic annotation. WebAnno is based on Java and SQL database, so it has quite standard requirements, making it relatively easy to run and operate. The application allows for sharing different projects in one installation.

During the annotation process, the annotators had access to some information from other layers existing in the gold-standard subcorpus of the National Corpus of Polish (NKJP1M), namely: morphosyntactic tags and some selected surface syntactic groups that could be indicators of quantificational usage of an expression. The syntactic groups are limited only to adverbial groups (which could be A-quantifiers) and numeral groups (most likely an existential numeral quantifier). However, as we treated quantifiers primarily as semantic units, annotators were not bound to those distinctions from other non-semantic layers. They were even free to switch off that information from their view if they did not consider it useful.

One may observe that the three quantifier features, grammatical structure, quantificational force, and monotonicity are to at least a significant extent lexico-morphosyntactic properties. Hence, one may wonder whether we could have first created an exhaustive dictionary of the quantifiers and only later assign the properties to the identified lexical quantifiers. We have

decided to create a quantifier dictionary and annotate the features in parallel because we did not want to assume that no two homonymous quantifying expressions can have different feature values depending on the context. Also, identifying quantifiers only based on dictionary entries is doubtful as some words may serve both as quantifiers and non-quantifiers depending on context. Consider two examples below: in (16) *większość* means majority as in parliamentary majority and is not a quantifier. In (17), however, *większość* 'most' is one of the usual proportional quantifiers:

- (16) W Sejmie jest większość dla podjęcia takiej uchwały.
 In parliament there is majority for adopting such resolution.
 There is a majority in the parliament for adopting such a resolution.
- (17) Większość mieszkańców Czeczenii sceptycznie odnosi się do szczerości Putina.
 Most inhabitants Chechnya skeptically refer to sincerity Putin's.
 Most people in Chechnya are skeptical about Putin's sincerity.

Annotation

1 Również wczoraj ogień całkowicie strawił drewniany domek jednorodzinny stojący przy ul. Mrągowskiej w podbydgoskim Myśliczku.

2 W momencie wybuchu pożaru w budynku nie było nikogo z domowników.

3 Jak nas poinformował oficer dyżurny KW PSP w Bydgoszczy, w budynku spłonęły najprawdopodobniej dwa psy.

4 Trwa ustalanie przyczyn wybuchu ognia, prawdopodobnie było to jednak zwarcie instalacji elektrycznej.

Invisible annotations on other pages

1_aleksanderl

1 Również wczoraj ogień całkowicie strawił drewniany domek jednorodzinny stojący przy ul. Mrągowskiej w podbydgoskim Myśliczku.

2 W momencie wybuchu pożaru w budynku nie było nikogo z domowników.

3 Jak nas poinformował oficer dyżurny KW PSP w Bydgoszczy, w budynku spłonęły najprawdopodobniej dwa psy.

4 Trwa ustalanie przyczyn wybuchu ognia, prawdopodobnie było to jednak zwarcie instalacji elektrycznej.

2_sebastianz

1 Również wczoraj ogień całkowicie strawił drewniany domek jednorodzinny stojący przy ul. Mrągowskiej w podbydgoskim Myśliczku.

2 W momencie wybuchu pożaru w budynku nie było nikogo z domowników.

3 Jak nas poinformował oficer dyżurny KW PSP w Bydgoszczy, w budynku spłonęły najprawdopodobniej dwa psy.

Figure 1. An example of adjudication of collision between two annotators as seen in the Webanno application. The screen shows the same sample consisting of four sentences: the top part is the adjudicator's version and the remaining two are versions of the annotators. As it could be seen, the annotators agree on everything but one feature of the quantifier *nikogo z* 'none of' in the second sentence (see example (6)): the second annotator decided that it is universal rather than existential. The second annotator is wrong and the adjudicator chose the version of the first annotator.

5. Inter-annotator Agreement

As it was mentioned above generalized quantifier expressions may be sparse textwise. For that reason, our approach was focused primarily on identifying all text units that could be interpreted as quantifiers and correct or reduce the redundant ones during the process of adjudication. Selecting two groups of annotators based on their background and experience was motivated by this goal, even though it resulted in a significant decrease in the inter-annotator agreement (IAA) rate.

As it is always the case in manual annotation a significant part of inconsistencies between annotators were simple mistakes, overlooks, and misclicks. Among more interesting examples of inconsistencies were quantifiers such as *żaden* 'none', *nikt* 'nobody', *nic* 'nothing', *nigdy* 'never', *nigdzie* 'nowhere' which quite often by at least some annotators were marked as universal rather than existential. From the cognitive perspective, this seems quite natural that expressions semantically informing of the non-existence of entities of some sort are actually existential, even though such examples were given in the annotation manual. Of course, we are not drawing any conclusions about the cognitive aspects of such quantifiers, however, we consider this specific example interesting.

Another typical example of systematic mistakes in the annotations are quantifiers such as *większość* 'most' and similar, which are monotonic on one argument and non-monotonic on the other. Some annotators intuitively and unconsciously assume that the quantifier needs to be either monotonic or non-monotonic on both arguments.

To calculate IAA we used Cohen's Kappa as a standard and most widely used measure. However, there are at least two problems to keep in mind. Firstly, quantifiers may be either single- or multi-token expressions so the annotator needs to identify boundaries of such expressions — two annotators may generally agree on the given quantifier and its features but disagree over one token belonging (or not) to the expression. Secondly, quantifiers are sparse, which means the tokens that are not annotated vastly outnumber those that need to

be annotated, which artificially increases the score. Thus, for the purpose of evaluation, we have calculated Cohen’s Kappa for the token level and only for those tokens which were annotated as belonging to quantificational expressions.

In the process of annotation 36,887 tokens were identified as belonging to such expressions by at least one annotator and only 23,284 (63,1%) simultaneously by both annotators. This strongly impairs the IAA results and suggests that the task of identifying quantifiers was relatively difficult. As stated above, low IAA for quantifiers identification was expected and to some extent influenced by the decision of recruiting annotators with diverse backgrounds and experience. However, if only tokens annotated by both annotators are taken into account it is possible to additionally tell which features are harder to specify after a quantifier is identified. Table 1 presents Cohen’s Kappa for four features described in section 3. As it could be seen, the IAA score is relatively high for quantifier type (D vs. A), with subtype (universal vs. existential vs. proportional) of a significantly lower rate and both left and right monotonicity with almost identically low rates. The ranking is again expected since the decision between A and D types is binary and to some extent could be induced based on surface syntactic features, so the task is relatively easier than deciding on other features. On the other hand, monotonicity itself is a complicated notion, and diagnosing it is even more difficult outside of the context of an artificially prepared textbook example.

The IAA results are rather low compared to the results of annotation for well-established NLP tasks such as named entities. However, our approach was experimental, and in the process of annotation, we intentionally favored recall over precision. Thus a large amount of work was also done in the process of adjudication.

Type (D or A)	0.90
Subtype (existential, universal, or proportional)	0.76
left monotonicity	0.62
right monotonicity	0.63

Table 1: IAA (Cohen’s Kappa) for separate features of quantifiers based on tokens annotated by both annotators. Type refers to the distinction between A- and D-quantifiers. Subtype refers to a distinction between universal, existential, and proportional quantifiers. We separately looked into the monotonicity in the left (restriction) and right (scope) arguments.

6. Querying the Corpus on the Web

The corpus is available on the web both as a separate layer of annotation together with the whole NKJP1M indexed in the corpus search engine and as an XML source tarball for processing in other projects, see: <http://kwantyfikatory.nlp.ipipan.waw.pl/>. The corpus is indexed using MTAS (Brouwer et al., 2017), a multi-tier annotation search engine that allows for indexing multiple layers of annotation. The quantifier layer, as well as all previously existing annotation layers, are all accessible from the Corpus Query Language, which enables searching for alignments between grammatical and quantificational layers. The scripts and data reported in the paper can also be found there and will be made available upon the publication of the paper.

The query language is an extension of the annotation layers previously existing in the corpus consisting of morphosyntax, named entities, syntactic groups, and some limited word sense disambiguation layers. The quantifier layer consists of single- and multiword elements together with their features encoded in the tagset. Quantifier units may be queried with a `<q />` element, optionally enriched by feature values combined in a positional tag. For example, a query:

```
<q="D:prop:nmon:inc" />
```

will search for all D-type proportional quantifiers, which are left non-monotone and right upward monotone. The majority of hits, in this case, will return instances of quantifiers such as *many*, *both*, *most*, or similar. Regular expressions are accepted in the query as well, allowing for queries with some features unspecified (e.g. `<q=".*:exst:.*" />` for all existential quantifiers). It is also possible to query simultaneously different layers of annotation, which allows for restricting the results to quantifiers containing specific words (e.g., `<q />` containing `[base="żaden"]` for quantifiers containing a word *żaden* 'no'), containing specific part of speech (e.g. `<q />` containing `[pos="conj"]` for quantifiers containing conjunctions) or occurring in specific phrases (e.g., `<q />` `fullyalignedwith <g="NumG" />` for quantifiers that are numeral groups on the syntactic level). Basically, MTAS allows for constructing queries compliant with the CQL standard.

7. Basic statistics

The NKJP1M corpus consists of 18,484 short samples (40-60 words long each, up to full sentence). In 11,606 (62.79%), at least one quantifier was annotated. In total, 21,938 quantificational expressions were annotated, which is 1,19 on average in each sample.

As expected, see Table 1, one of the most numerous groups among the quantifiers is unmodified numerals constituting 29.82% of all units. D-quantifiers, including the unmodified numerals (27%), are ten times more frequent than A-quantifiers (19801 to 2137). Existential quantifiers (7461, 34.01%) are more frequent than universal ones (3902, 17.79%), which are again slightly more frequent than proportional (4032, 18.38%). Next to typical examples for the class of proportional quantifiers, this count includes, as discussed above, *oba* and *obydwa* 'both' (8,26% of all proportional count). Moreover, percentage expressions of the form "n %" or quantifiers including the sequence "n%," e.g., "more than n%" are counted under the proportional label (379 occurrences). These frequency numbers are consistent with the semantic complexity predictions mentioned in the introduction (Szymanik and Thorne, 2017).

Feature query	Feature count	%
A	2137	9.74
D	19801	90.26
exst	7461	34.01
univ	3902	17.79
num	6543	29.82
prop	4032	18.38
left inc	2747	12.52
left dec	6363	29.0
left nmon	12828	58.47
right inc	9312	42.45
right dec	2882	13.14
right nmon	9744	44.42
pos	21559	98.27
com	208	0.95
sup	171	0.78

Table 2: The table presents the number of quantifiers in the corpus for each value of annotated categories. Counts in each category sum up to 21,055 (100%).

	D	A	univ	exst	num	prop	left_inc	left_dec	left_nmon	right_inc	right_dec	right_nmon
D	X	X										
A	X	X										
univ	3357	545	X	X	X	X						
exst	6507	954	X	X	X	X						
num	6543	0	X	X	X	X						
prop	3394	638	X	X	X	X						
left_inc	2359	388	99	2648	0	0	X	X	X			
left_dec	5439	924	3803	2560	0	0	X	X	X			
left_nmon	12003	825	0	2253	6543	4032	X	X	X			
right_inc	7842	147 0	3804	2745	0	2763	2648	3834	2830	X	X	X
right_dec	2365	517	98	2532	0	252	99	2529	254	X	X	X
right_nmon	9594	150	0	2184	6543	1017	0	0	9744	X	X	X

Table 3: The table presents the number of quantifiers in the corpus for each pairwise combination of annotation values.

8. Corpus and Quantifier Theory

One of the crucial achievements of formal semantics was to formulate linguistic universals for the domains of function words, with quantifiers being a prime example in the literature (Barwise and Cooper 1981). Over the years, we have seen intensive research efforts in linguistics and cognitive science to assess the proposed universals' empirical adequacy and find explanations for their existence (see, e.g., Steinert-Threlkeld and Szymanik, 2020a). This research program's major bottleneck is the lack of cross-linguistic quantitative data that could help with the theory testing and development. Such data often exists for other semantic domains, e.g., color terms (Berlin and Kay, 1969) or kinship terms (Murdock, 1970), supporting advancements in the field (Kemp & Regier, 2012, Steinert-Threlkeld & Szymanik, 2020b). We believe that gathering quantitative cross-linguistic data on function words, like quantifiers, is crucial to understand language and cognition better. Therefore, our biggest hope is that the detailed description of the corpus building process, presented in the paper, will motivate similar work and be a reference to replicate the process in other languages.

What does our data tell us already about the quantitative distribution of quantifier features? A review of the annotation regarding right monotonicity may be at first sight surprising for researchers working on quantifiers. We have tagged 44.42% of all quantifiers as non-monotone in the right argument (right nmon, see Table 2). However, the quantifier literature strongly suggests that right monotonicity is a cross-linguistic universal among monomorphemic quantifiers (Barwise & Cooper, 1981). So let us break down the class of right non-monotone quantifiers in Polish into further categories: first of all, 67% of all occurrences are, in fact, numerals. Hence, they are either non-monomorphemic quantifiers of the form “exactly n” or just bare numerals of the form “n,” which arguably should be interpreted semantically in a monotone way as “at least n.” For instance,

(19) W sumie pełnił funkcję prezydenta przez 5 lat i 214 dni.

In total performed function president for 5 years and 214 days.

In total he performed the presidential function for 5 years and 214 days.

Another interesting subgroup, existential right nonmonotone quantifiers (22%), consists of very specific to Slavic languages quantifiers such as kilka, kilkanaście, kilkadziesiąt, kilkaset meaning ‘more than X and less than Y’ (e.g., kilkanaście means ‘between ten and twenty’) and their A-type adverbial counterparts (kilkakrotnie, kilkunastokrotnie). For example,

(20) Po tym wydarzeniu przez kilka dni przebywał w szpitalu.

After this event for between 1-10 days stayed in hospital.

After this event he stayed between one and 10 days in the hospital.

Arguably, such quantifiers also cannot be labeled as morphologically simple. Moreover, they are semantically equivalent to the conjunction of monotone quantifiers. They are so-called connected quantifiers. The remaining 10% consists of complex proportional quantifiers including the phrase “n%,” which got assigned the exact interpretation, exactly “n%” (out of 379 occurrences of such quantifiers, only 51 are monotone). For instance, see the second part of sentence (9) repeated below:

(21) ... a tylko 9 proc. to mieszkańcy wsi.

... and only 9 percent are inhabitants countryside.

... and only 9 percent are inhabitants of the countryside.

Hence, even though Polish is not a counterexample to the current universal formulation, the data shows that non-monotone complex quantifiers are very frequent textwise.

Monotonicity also plays a crucial role in psycholinguistic research. The experimental studies have repeatedly demonstrated that downward monotone quantifiers are more difficult to process than upward monotone ones. Subjects need more time to read sentences with downward monotone quantifiers and make more errors when asked to evaluate their truth values (see, e.g., Szymanik 2016 for an overview). One possible explanation of this so-called monotonicity effect is that it may be due to the relative frequencies (Degen and Tanenhaus, 2019). And indeed, in our corpus, the number of right downward monotone quantifiers is significantly smaller than the number of right upward monotone quantifiers (see Table 2).

9. Outlook

The first step in future work with the annotated corpus is an insightful analysis of the annotated units concerning the possible extensions of quantifier description. The three categories considered in our project are by no means exhaustive, and many other possible features of quantifiers could be added in the extended annotation. The list of categories can be extended, depending on the research goals. For instance, among existential quantifiers, one may wish to distinguish value judgments, e.g., “Enough members attended to constitute a quorum” (Keenan and Paperno, 2012) and among non-monotone quantifiers, one may want to distinguish connected quantifiers (conjunction of monotone quantifiers), e.g., “between 5 and 7” or “kilkanaście” (Chemla et al., 2019). Furthermore, building on already determined quantifier features, one may want to focus, for instance, on morphosyntactically complex quantifiers, like already mentioned modifications but also Boolean combinations, exception phrases (*all but students*), bounding phrases (*twice a day*), or partitive constructions (*most of the*) (Keenan and Paperno, 2017). Another direction would be to try to disambiguate and count various readings of quantifiers, for example, proportional and cardinal readings of quantifiers such as *many* or *few*. The tagging system could also be extended by other quantifier properties known in the literature like universe independence also known as extensionality (Peters and Westerståhl, 2006). Some of those extensions may be carried out automatically or semi-automatically. Another possible extension of the annotation is to include the comparison type of quantifiers. Each quantifier can be either positive, comparative, or superlative. Modified numerals come in two, semantically equivalent, flavors: comparative, e.g., *more than*, *fewer than*, and superlative, e.g., *at least*, *at most*. Geurts et al., (2010) have provided evidence that superlative quantifiers are harder

to process for humans than comparative quantifiers. Thus, as in the case of monotonicity, it may be interesting to compare the frequency of the two types of modified numerals.

In parallel with our annotation project, there has been recently an effort to establish an ISO standard annotation scheme for quantification phenomena in natural language as part of the ISO Semantic Annotation Framework (ISO 24617) (Bunt, 2020). The developments of the standard are still very much in the preliminary stage. Most importantly, it still needs to be validated in a manual and automatic annotation. The annotation scheme proposed in the standard is highly complex in its current form, making it too difficult to use for the annotators. However, if the system will become validated and supported by training and annotation tools, it would be interesting to test it on our corpus.

The manually annotated data will also serve as a training corpus for a machine learning classifier aimed at the automatic semantic annotation of quantifiers in large corpora. Based on the annotation, we plan to carry out an extensive corpus-based analysis of quantifiers distribution in Polish based on the standard balanced and representative 300M tokens large corpus of modern Polish. One natural direction would be repeating the research on semantic complexity conducted by Szymanik and Thorne (2017). The biggest weakness of their analysis was the restriction to 36 most common quantifiers. Using our corpus, we could have much broader coverage, approximating all quantifier expressions in Polish. Therefore any statistical generalization about the influence of various semantic factors on linguistic distribution would be more robust. Also, such analysis would be based on a typologically different language than English. Furthermore, an additional aspect of the text genre could be taken into account.

Acknowledgments

The authors have been supported by the Polish National Science Centre grant number 2017/25/B/HS1/02911. The authors are also grateful to Dorota Komosińska for help with developing the annotation tool and Maciej Ogrodniczuk for support in running the project.

Bibliographical References

Bach, E., Jelinek, E., Kratzer, A., and Partee, B. H. (eds.) (1995). Quantification in Natural Languages, volume 54 of Studies in Linguistics and Philosophy. Springer.

Barwise, J. and Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4, pp. 159–219.

Baroni, M., Bernardini, S., Ferraresi, A., and Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43 (3), pp. 209–226.

Berlin, B. and Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley and Los Angeles. University of California Press.

Brouwer, M., Brugman, H., and Kemps-Snijders, M. (2017). MTAS: A Solr/Lucene based Multi Tier Annotation Search solution. In *Selected papers from the CLARIN Annual Conference 2016, Aix-en-Provence, 26–28 October 2016, CLARIN Common Language Resources and Technology Infrastructure*, 136, pp. 19-37. Linköping University Electronic Press ,Linköpings universitet.

Bunt, H. (2020). Annotation of Quantification: The Current State of {ISO} 24617-12. In *the Proceedings of 16th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*. European Language Resources Association.

Chemla, E., Buccola, B., and Dautriche, I. (2019). Connecting content and logical words. *Journal of Semantics*, 36(3) pp. 531–547.

Degen, J. and Tanenhaus, M. K. (2019). Constraint-based pragmatic processing. In *The Oxford Handbook of Experimental Semantics and Pragmatics*.

Deschamps, I., Agmon, G., Loewenstein, Y., and Grodzinsky, Y. (2015). The processing of polar quantifiers, and numerosity perception. *Cognition*, 143, pp. 115-128.

Eckart de Castilho, R., Mújdricza-Maydt, É., Yimam, S. M., Hartmann, S., Gurevych, I., Frank, A., and Biemann, C. (2016). A web-based tool for the integrated annotation of semantic and syntactic structures. In Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH), pages 76-84, Osaka, Japan, December. The COLING 2016 Organizing Committee.

Geurts, B., Katsos, N., Cummins, C., Moons, J., and Noordman, L. (2010). Scalar quantifiers: Logic, acquisition, and processing. *Language and Cognitive Processes*, 25 (1), pp. 130-148.

Keenan, E. and Paperno, D. (2012). *Handbook of Quantifiers in Natural Language. Studies in Linguistics and Philosophy.* Springer Netherlands.

Kemp, C. and Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, 336, pp. 1049-1054.

Ladusaw, W. (1979). *Polarity Sensitivity as Inherent Scope Relations.* PhD thesis, University of Texas.

Murdock, G. P. (1970). Kin term patterns and their distribution. *Ethnology*, 9 (2), pp. 165-208.

Paperno, D. and Keenan, E. (2017). *Handbook of Quantifiers in Natural Language.* Number, vol. 2, in *Studies in Linguistics and Philosophy.* Springer International Publishing.

Paperno, D. (2012). Quantification in Standard Russian. In *Handbook of Quantifiers in Natural Language* (Keenan and Paperno, 2012).

Partee, B. H. (1988). Many quantifiers. In *Proceedings of the Eastern States Conference on Linguistics* (Vol. 5, pp. 383-402).

Partee, B. H., (1995). Quantificational Structures and Compositionality, pp. 541-601. Springer Netherlands, Dordrecht.

Peters, S. and Westerståhl, D. (2006). Quantifiers in Language and Logic. Clarendon Press, Oxford.

Przepiórkowski, A., Bańko, M., Górski, R. L., and Lewandowska-Tomaszczyk B. (eds.) (2012). Narodowy Korpus Języka Polskiego. Warszawa.

Steinert-Threlkeld S. and Szymanik J. (2020a). Learnability and Semantic Universals, Semantics & Pragmatics, Volume 12, Issue 4, 2020.

Steinert-Threlkeld S. and Szymanik J. (2020b). Ease of Learning Explains Semantic Universals, Cognition, vol 195.

Szymanik, J. and Thorne, C. (2017). Exploring the relation of semantic complexity and quantifier distribution in large corpora. Language Sciences 60, pp. 80-93.

Szymanik, J. (2016). Quantifiers and Cognition. Logical and Computational Perspectives. Studies in Linguistics and Philosophy. Springer.