Saarland University
Department of Computer Science
Winter term 2020/2021
Seminar Commonsense knowledge extraction and consolidation
Lecturer: Simon Razniewski

Seminar Paper

Modelling Generics

Name:	Priyanka Upadhyay
Matriculation number:	2581714
Field of study:	Master of Science in Data Science and Artificial Intelligence, (semester 3)
Email:	s8prupad@stud.uni-saarland.de
Date:	January 14, 2021

I hereby confirm that I have written the seminar paper with the title **Modelling Generics** in the seminar **Commonsense knowledge extraction and consolidation** (lecturer: **Simon Razniewski**) on my own and that I have not used any other media or materials than the ones referred to in this seminar paper.
I further confirm that I did not submit this or a very similar seminar paper in another seminar.

Saarbrücken, January 14, 2021 Priyanka Upadhyay

Contents

1 Abstract

Potatoes contain vitamin C, amino acids, protein and thiamine, examplepresses a true generalization about potatoes. John smokes a cigar after dinner, in generic sense is expresses a regularity in John's behavior after dinner, it could be true or could be false. This realist conviction inspires the theory of genericity which we will present in this report. Genericity is clearly a feature of the whole sentences (or clause), rather than of any one Noun Phrase in it. It is a whole generic sentence that expresses regularities which transcend particular facts. We will call these types of sentences Generic sentences or characterizing sentences, as they express generalization. These sentences are opposed to particular sentences, which express statements about particular event. The above 2 sentences can be true at all, their truth somehow depend on the nutritional value of the particular potatoes and on John's behavior on particular occasions after dinner. But this is the puzzling things about generics: their truth conditions connect them at best only very loosely with particular facts about the world. This tolerance of exceptions has been for decades in linguistics, philosophy and artificial Intelligence to provide rigorous account of generic meaning and of modes of argument to which generics sentences give rise. This paper will provide a deep understanding of generics sentences and use of them in NLP field through GENERICSKB. We present a new resource for the NLP community, namely a large knowledge base of generic statements, collected from multiple corpora.

2 Introduction

While Deep Learning systems have achieved remarkable performance trained on general text, NLP researchers frequently seek out additional repositories of general/commonsense knowledge to boost performance further. However, there are only a limited number of repositories currently available, with ConceptNet and WordNet being popular choices. The resource, called GENERICSKB, is the first to contain naturally occurring generic sentences, as opposed to extracted or crowdsourced triples, and thus is rich in high-quality, general, semantically complete statements.

We have also create GENERICSKB-BEST with 1M+sentences, which contain the best-quality generics from GENERICSKB plus selected, synthesized generics from WordNet and ConceptNet.

Our goal isn't to make a new model, but to work out how an existing model's performance changes when the GENERICSKB corpus replaces a bigger corpus for these tasks. We find that GENERICSKB can sometimes produce higher question-answering scores, and always produced better quality explanations, this means that GENERICSKB may have value for other NLP tasks also, either standalone or as an extra source of cognition to assist train models. Finally, independent of deep learning, GENERICSKB maybe a valuable resource for those studying generics and their semantics in linguistics.

As a result, we will show the use of GENERICSKB for two tasks, namely Question-answering , (using openbook QA Dataset [10]) and Explanation generation (using the QASC dataset [7]).

The remainder of the seminar paper is structured as follows. In Section 3.1, information about Generics, What are Generics Statements are explained. In Section 4, Dataset is explained and if there is any limitation of this Dataset In Section 5, We provide the Approach to generate our GENERICSKB - see introduction part and approach part In Section 6, we will talk about the Challenges which author and I faced. In Section 7, Evaluation from the author and my side are included there. In Section 8, Conclusion/Summary of the report you can find.

1. Example generics about "tree" in GENERICSKB

Trees are perennial plants that have long woody trunks.
Trees are woody plants which continue growing until they die.
Most **trees** add one new ring for each year of growth.
Trees produce oxygen by absorbing carbon dioxide from the air.
Trees are large, generally single-stemmed, woody plants.
Trees live in cavities or hollows.
Trees grow using photosynthesis, absorbing carbon dioxide and releasing oxygen.

2. An example entry, including metadata

Term: tree
Sent: Most trees add one new ring for each year of growth.
Quantifier: Most
Score: 0.35
Before: ...Notice how the extractor holds the core as it is removed from inside the hollow center of the bit. Tree cores are extracted with an increment borer.
After: The width of each annual ring may be a reflection of forest stand dynamics. Dendrochronology, the study of annual growth rings, has become prominent in ecology...

Figure 1: Example of generic statements in GENERICSKB plus one showing associated metadata.[1].

3 Generics

3.1 What is Generic?

[14]History of both Philosophy of Language and Linguistic gave us two distinct phenomena that have been referred as Genericity.

- **First Phenomena**: Generic Noun phrase (NP) in my study does not state to an ordinary individual or object but instead it refers to a Kind/group. These sentences are called **Kind referring NPs or Generics NPs**.
Ex: The potato was first cultivated in South America.
Ex: Potatoes were introduced into Ireland by the end of the 17th century.

- **Second Phenomena**: Genericity is clearly a feature of the whole sentences (or clause), rather than of any one Noun Phrase in it. It is a whole generic sentence that expresses **regularities which transcend particular facts**.
Ex: A potato contains vitamin C, amino acids, protein and thiamine.

We will call these sentences **Generic sentences or characterizing sentences**as they express generalization. These sentences are opposed to particular sentences, which express statements about particular event. Sometimes both Phenomena kind referring NPs and Generics Sentences can occur combined. Ex: The potato is highly digestible.

We want to focus on Generics statement in this research and it is very important that these 2 phenomena are kept apart for studies. One reason is that Generics sentences, in general allow for exceptions, whereas Kind refereeing NPs make claim for every object of a certain sort.

Ex: ConceptNet[9]:
1. Lion is an Animal.
2. Lion is capable of roar.
3. Lion sounds scary
4. Lion is a big cat.

WordNet[11]:
Lion, king of beasts, Panthera leo (large gregarious predatory feline of Africa and India having a tawny coat with a shaggy mane in the male)

GENERICSKB:
1. Lions have manes.

3.2 Generics as per old literature

Besides the GENERICSKB[1], there are few old research which gave the sense of a statement to be Generic and I am going to present few of them.

3.2.1 Hale et al., 1997

[6] This book has given quite clear meaning of Generics and Genericity,

There is a spread of sentences that, speaking intuitively, we are able to use to speak about kinds. The clearest examples are sentences that predicate properties that only kinds can have (what we may call kind-restricted predicates).

(1) Ravens are widespread.
(2) Dodos are extinct.
(3) Diamonds are rare.

Individual ravens can't be widespread; individual dodos can only be dead; and individual diamonds are all, by their very nature, unique.1 Since (1)–(3) concern kinds – **genera** – the claims themselves have come to be called generics.
In addition to describing a form using kind-restricted predicates, we will also characterize it, or perhaps characterize its members **qua** members of that sort, by using properties that individuals can have. we will call these latter generics **characterizing generics**. (4)–(12) are some true characterizing generics.

(4) Ravens are black.
(5) Tigers have stripes.
(6) Coke bottles have short necks.
(7) Bishops move along diagonals.
(8) Lions have manes.
(9) Lions give birth to live young.
(10) Lions have four legs.
(11) Sea turtles are long-lived.
(12) Barns are red.

We can contrast them with (13)–(17), which are false.
(13) Ravens are white.
(14) Prime numbers are odd.
(15) Supreme Court Justices have odd Social Security Numbers.
(16) Green bottles have short necks.

(17) Sea turtles die young.

As these examples indicate, plenty of our knowledge of the world is most naturally expressed using generic sentences. Generic sentences are a crucial component of the language young children hear as they mature **("motherese")**. For that reason, a semantic theory for natural language must have an account of generics.
These examples share several more specific features – beyond the intuitive sense that they are closely connected to kinds – that have made them of interest to philosophers and linguists alike, and these specific features support the intuitive sense that generics are particularly closely concerned with kinds.

3.2.2 Greg N. Carlson, 1995

[2]This paper has provided a very clear structure of Semantic Composition of English Generic Sentences.Notionally, a generic sentence is one expressing a regularity, as opposition an instance from which one infers a regularity. For example, the generalization "The sun rises within the east" expresses a regularity, while "The sun rose this morning within the east" expresses an instance from which, together with other such instances, one infers a regularity. Epistemologically, a generic sentence is one expressing a truth (or falsehood) the truth value of which cannot, in general, be ascertained solely with reference to any particular localized time. as an example, this tense sentence "Dogs bark" is true, while at the current time there is also no dogs barking. Or, take the assertion "Bears hibernate" said in midsummer; to test on that one must stay up for winter. In contrast are the specific instances, which generally may be "locally determined". If I say "Jenny is watching TV within the den," this can be something I can, a minimum of in principle, check on straight away, and other times become irrelevant to the reality or falsity of this assertion. Or if I assert that Bruno is in hibernation, I check thereon now; I don't wait until midwinter. Linguistically, generic sentences don't have any formal distinguishing features, at least in English and plenty of (though not all) other languages (see Dahl (1985) for an overview).

3.3 Generic Vs Non-Generic

In this particular section, we are going to cover the following points:
- How easy it is to say that a statement in Generic?
- What are the clear distinction of Generic Vs Non-Generic?

- 27 Rules of Patterns for Identifying Generics

[2]Surface syntax, morphology, and phonology don't distinguish generics from non-generics. They do, however, have some distinguishing semantic features: generic sentences are (i) stative sentences (ii) supported lexically non-stative predicates, and (iii) they're intensional and (by all appearances) non-monotonic. as an example, the aspectual category of "fetch newspapers" is non-stative (a process), yet the sentence "Dogs fetch papers" passes all the quality tests for stativity.

3.3.1 Generic Sentences

[12]We will give a brief idea to say if the sentence is generic.

- Do not report specific or isolated facts, but express a kind of general property
Like: a regularity summarizing groups of particular episodes or events or facts or states of affairs

- Characteristic of or relating to a class or group of things, not specific
Ex: Chèvre is a generic term for all goat's milk cheese.

- Much of our commonsense knowledge of the world is expressed by generic sentences
Ex: Potatoes contain vitamin C
Ex: The lion has a mane

- Not only distinct from individual or particular predications, but also from explicit quantificational sentences
Ex: Each potato contains vitamin C
Ex: Most potatoes contain vitamin C
Ex: All potatoes from Alberta taste good

Exceptions to Generic Sentences:

"Exception Tolerating"
Ex: – Some lions do not have manes
Ex: – Some potatoes are indigestible

"Two Bad Attitudes"
1. Generics are properly speaking false, but are tolerated because they're

‘close enough for practical purposes
2. Generics are neither true nor false; they are ‘rules to measure by’ ‘ways to draw inferences’, etc

"Demur from the first attitude"
Most of our everyday, commonsense knowledge of the world is encoded in generic sentences So, it’s pointless to mention that they’re merely acceptable and not really true. Further, if correct, we might expect that “the fewer exceptions, the more acceptable”. But this can be wrong.

"Demur from the second attitude"
- Denies that Snow is white is either true or false!!
- Denies that our information about the world is knowledge, but instead claims it to be “how to direct our actions and inferences”.

But would deny generics can be embedded?
Ex: Usually, if a person smokes after dinner, he also drinks brandy before bed.
Ex: People who work late nights do not wake up early.

3.3.2 The Generic/Nongeneric Distinction

[3]For adult it may be easy to find a distinction but a very important aspect we may forget that Generic/Nongeneric Distinction influences How Children Interpret new information about Social Others and I want to introduce it briefly. [3]The paper investigate how the distinction between Generic sentences (e.g., "Boys are good at math") and Nongeneric sentences (e.g., "Johnny is good at math") shapes children’s social cognition.

General Discussion Generic sentences like "Boys are good at math" convey broad generalizations about entire categories of individuals. Given their prevalence in speech to children and their robust semantic properties (e.g., their resistance to counterexam- ples), generics constitute a very important means of learning about others. What children learn from a generic sentence, however, might not be limited to a straightforward mapping between a property (e.g., being good at math) and a broad referent set (e.g., boys). The hypothesis in the study was that generic language also conveys, more covertly, a specific perspective on how this mapping came to be, inducing children to think about the relevant prop- erty as emerging naturally from an internal source.

The summary described that the studies add to this rich literature by suggesting that the semantic distinction between generic and nongeneric sentences carries more meaning than one might at first suspect. If we add to the scope of the properties conveyed about generic and nongeneric sentences, they also differ in their implications about the source of their findings described within the present studies augment this rich literature by suggesting that the semantic distinction between generic and nongen- eric sentences carries more meaning than one might initially suspect. additionally to the scope of the prop- erties conveyed, generic and nongeneric sentences differ in their implications about the source of those properties and thus in their implications about the type of properties they are-essential, stable fea- tures of their referents that emerge effortlessly from an indoor source versus superficial, temporary fea- tures that emerge as a results of external intervention or sustained effort. However, our studies also show that children won't automatically essential- ize any piece of knowledge that's conveyed in generic form. Rather, they actively integrate this linguistic cue with their theoretical knowledge about the relevant categories and properties, and are thus likely to hit an expensive, nuanced inter- pretation of the facts they learn.

3.3.3 Patterns for Identifying Generics (27 Rules)

[1]The following 27 rules are used to identify generic sentences, as well as help filter out those which are likely contextual, gibberish, or otherwise not stand-alone. Some rules use spaCy features for processing. To be retained, each sentence must pass the following tests:

is-short-enough: Length of the sentence

$$\leq 100.$$

starts-with-capital: The first character is an upper-case character.

ends-with-period: The last character is a period.

has-at-least-one-token: The sentence contains at least one spaCy token.

has-no-bad-first-word: The first word is not in a list of bad-first-words (determiners, etc.)

has-no-bad-words: The sentence does not contain words in a badword list (e.g., copyright, licence, ...)

has-no-bad-pronouns: The sentence does not contain personal pronouns (he, she, ...)

has-no-negations: The sentence does not contain negations.

has-no-modals: The sentence does not contain modals (“would”, “should”,...).

first-word-is-not-verb: The first word of the sentence is not a verb.

first-word-is-not-conjunction: The first word is not a conjunction.

look-for-positive-quantifier-at-first-word: If the first word is a positive quantifier (“all”, “some”), note the quantifier and repeat the filter using the sentence without the quantifier.

has-acceptable-past-participle-root: The root verb is in the present passive, or is not a past participle.

noun-exists-before-root: There is a ’NOUN’ token before the root.

key-concept-head-pos-tags-not-contradictedby- wordnet: If WordNet disagrees about the POS of the key concept head, filter out this sentence.

has-no-digits: The sentence has no digits.

all-propn-exist-in-wordnet: All PROPN tokens exist in WordNet.

all-propn-have-acceptable-ne-labels: Any PROPN tokens have one of the following ent type values: ’EVENT’, ’GPE’, ’LANGUAGE’, ’LAW’, ’LOC’, ’WORK OF ART’.

(These above acceptable values were decided by the corresponding top level rules.)
and must not pass these below tests:

scr.dot dot in sentence: There is ’..’ in the sentence.

scr.www in sentence: There is ’www’ in the sentence.

scr.com in sentence: There is ’.com’ in the sentence.

scr.many hyphens in sentence: The number of hyphens in the sentence is

$$\geq 2.$$

scr.sentence does not end with period: The sentence does not end with a period.

remove-non-verb-roots: Remove any sentences with non-verbal roots (e.g., “A large tree.”).

remove-present-participle-roots: Reject sentences whose root verb is a present participle ("sitting",...).

remove-first-word-roots: Reject sentences with a root that corresponds to the first word.

remove-past-tense-roots: Reject sentences with any past tense roots ("ate",...).

4 Dataset

[1]Statements in GENERICSKB were pulled from over 1.7 billion sentences from three corpora: Waterloo[5], SimpleWiki(simple.wikipedia.org), ARC[4].

Paper also talks about the GENERICSKB-BEST which containing the best-quality generics in GENERICSKB plus selected, synthesized generics from WordNet[8] and ConceptNet[13].

As says that generic statement is one that makes a blanket statement about the members of a category, Ex: "Tigers are striped. and because of they apply to too many entities, they are particularly important for reasoning. Our primary goal is to collect Generics rather than interpret them. We hope that our resource can contribute to study of Generic statement semantics. Several repositories of general knowledge are available already, but with different characteristics and coverage to GENERICSKB. ConceptNet[13] is perhaps the most used repository (containing approximately 1M English triples excluding RelatedTo, Synonym, and Lexical FormOf links, or 34M triples total). ConceptNet triples can be rendered as short generics and can be used as dataset in GENERICSKB, thus covering just simple (typically three word) generic statements about 28 relationships. Similarly, WordNet[8] repository taxonomic and meronymic links express short, specific relationships but leave most uncovered topics(compare with Figure 1). Triple stores acquired from open information extraction contain larger and less constrained collections of knowledge, but typically with low precision making it difficult to exploit them in practice. GENERICSKB thus fills a gap in this space, containing naturally occurring generic statements that an author considered salient enough to write down.

5 Approach

[1]Over 1.7 billions sentences from three corpora (check Figure 2) were selected to construct GENERICSKB.

Corpus	Size (# sentences) Original	Cleaned	Filtered
Waterloo	∼ 1.7B	∼ 500M	∼ 3.1M
SimpleWiki	∼ 900k	∼ 790k	∼ 13k
ARC	∼ 14M	∼ 6.2M	∼ 338k
GENERICSKB	∼ 1.7B	∼ 513M	∼ 3.4M

Figure 2: Corpus sizes at different steps of processing[1].

- The Waterloo corpus is English plain text with the size of 280GB. It was gathered by using a webcrawler in 2001 from .edu domains and then it was made available to public and was previously used in [5].
- SimpleWikipedia is a filtered scrape of SimpleWikipedia pages (simple.wikipedia.org).
- The ARC corpus is a collection of 14M science and general sentences, released as part of the ARC challenge [4].

GENERICSKB was then assembled in the following three steps with sentence from above 3 corpora. To collect statements from 3 corpora and to make it usable in GENERICSKB (1.7B Normal sentences were collected from 3 corpora and after these 3 steps the sentence which will have are 3.4M Generic sentences).

Step1: Cleaning: we first clean the source data,
Step2: Filtering to find likely generic sentences within normal sentences: We filter the sentences using linguistic 27 rules (see 3.3.3) to identify likely generics.
Step3: Scoring to find semantic generics: then we apply a BERT-based scoring step to distinguish generics that are meaningful on their own (avoiding generics with contextual meaning such as Meals are on the third floor).

At the end of the last step, we will get our resulting KB containing over 3.5 millions Generics statements, each including metadata about its topic, surrounding context, and a confidence measure.

We also create GENERICSKB-BEST (1M+ sentences), containing the best-quality generics in GENERICSKB plus selected, synthesized generics from WordNet and ConceptNet.

Here are the 3 steps:

5.1 Cleaning

3 source corpora are originated from web scrapes so it's quite common that they contain noise in various forms, like hyperlinks, blocks of code, non-English text, and emails. And to use this data, we'd like to wash it first with the subsequent process:

- Regular Expressions (occurring lexical properties of noise) to capture frequently.
- Sentence and token length heuristics to separate malformed sentences.
- Text cleanup using the Fixes Text For You (FTFY) using python library (to fix various encoding-related errors).
- Language Detection using spaCy to filter non-English text.

5.2 Filtering

We are visiting the 27 rules3.3.3 to identity the standalone/independent generics sentence and to reject other which doesn't follow the foundations. For example,

1. Sentences that start with a bare plural ("Dogs are...") are considered nearly as good candidates, while those starting with a determiner ("A man said...") or containing a present participle ("A bear is running...") are not considered pretty much as good generics and are rejected.
2. Sentences containing pronouns ("He said...") are likely to possess contextual rather than standalone meaning, and then also are rejected.

Please see the filtering rules sample in figure 3 and full list of rules is given within the section 3.3.3.
Given the size and redundancy of the initial corpus, these rules aim to filter the corpus aggressively to provide a group of high-quality candidates, rather than catch all possible standalone generics. At the tip of the method we are going to collect the likely generics sentences.

no-bad-first-word: Sentence does not start with a determiner ("a", "the",...) or selected other words.
remove-non-verb-roots: Remove if root is a non-verb
remove-present-participle-roots: Do not consider any present participle roots.
has-no-modals: Sentences containing modals ("could", "would", etc) are rejected
all-propn-exist-in-wordnet: All (normalized, non-stop) words are in WordNet's vocabulary

Figure 3: Sample of filtering rules[1].

5.3 Scoring

The scoring step is employed to search out semantic generics. A BERT classifier (used for Text classification in NLP) is applied on the dataset to score sentences by how well they describe a useful, general truth, in other ways to differentiate generic statements that are meaningful on their own. To build the BERT classifier, a random subset of size 10k (from 3.4 millions candidate generics) was labeled by crowd workers to require their suggestion and to record their expression as a useful, general truth about the world (with having 3 options: yes, no, unsure), guided by examples. Specifically, workers were asked to reject:

(1) Sentences that don't stand on their own,
Ex: Free parking is provided

(2) subjective and/or not useful statements,
Ex:Life is simply too serious, sometimes.

(3) Vague statements,
Ex:All cats are essentially cats.

(4) Statements about people and firms,
Ex:Apple makes plenty of iPhones.

(5) Facts that are incorrect in isolation,
Ex: All maps are hand-drawn.

BERT Classifier functioning:

- Each fact was annotated twice and scores (yes/unsure/no = 1/0.5/0) averaged.

- The joint probability of agreement (i.e., that both annotators agreed) was 70.1% (approximately 1/3 of the agreed annotations being "yes" and 2/3 "no"), and Cohen's Kappa coefficient () was 0.52 ("moderate agreement").

- Afterward, the Dataset was then split 70:10:20 into train:dev:test, and a BERT classifier fine-tuned on the training set.

- Each sentence is input simply as sentence and The output is pooled, then run through a linear layer which have two logits outputs representing the two classes (yes/no), followed by a softmax to get class probabilities. This classifier scored 83% on the held-out test set.

-The classifier was then made to score all 3.4 millions extracted generic sentences.

5.4 GENERICSKB-BEST

We used GENERICSKB[1], Conceptnet[13], Wordnet[8] , and Aristo TupleKB (at https://allenai.org/data/tuple-kb) as our source corpora to form GENERICSKB-BEST. GENERICSKB-BEST is comprising GENERICSKB generics with a score > 0.23 (By calibration, equivalent to an annotator score of 0.5, i.e., more likely good than bad).

The GENERICSKB-BEST is augmented with short generics synthesized from three other resources (ConceptNet (isa, hasPart, locatedAt, usedFor); WordNet (isa, hasPart); and also the Aristo TupleKB. For WordNet, we did use just the most frequent sense for every generic term.) for all the terms (generic categories) in GENERICSKBBEST. GENERICSKB-BEST contains 1,020,868 generics (774,621 from GENERICSKB plus 246,247 synthesized from Concept and Wordnet).

6 Challenges

The challenges we face into writing down the perfect rules to identify the generics is quite difficult and may some statement pass these rules but actually are not generics. To understand what is actually a generic statement and what semantics is provided, I needed to refer multiple resources but still at the end it provide some ambiguity and we may fail for some uncommon situation to identify a generics. GENERICSKB-BEST has solved this problem with 85% of accuracy (see section 7.3)

Corpus	Size	Score on OBQA (test)
QASC-17M	17M	0.660
GENERICSKB	3.4M	0.632
GENERICSKB-BEST	1M	**0.678**

Figure 4: Comparative performance of different corpora for answering OBQA questions[1].

7 Evaluation

We can see the purpose to build GENERICSKB while performing 2 experiments in NLP.

7.1 Question-Answering

We evaluate using GENERICSKB for a question-answering task, namely OpenbookQA [10] comparing it to using an alternative, large, publically available corpus (QASC-17M [7]). For both, we use the BERT-MCQ QA system [7]. Note that our goal is to judge the corpora, not the QA system. The results are shown in Figure 4, indicating that using the high-quality version GENERICSKB-BEST can, a minimum of during this case, lead to improved QA performance over using the first corpus, while it's a fraction of the scale.

7.2 Explanation Quality

We used GENERICSKBBEST to get explanations for a (given) answer and also the explanation may be a chain of two sentences (taken from the corpus).

Example: **What can cause a forest fire? storms because:**
Storms can produce lightning
AND Lightning can start fires

As we all know that good explanations mostly use generic sentences, which reflect the underlying formal structure of the reason. This suggests that a corpus of generics may help during this task. To test this hypothesis we used the QASC dataset with original QASC-17M corpus, and using GENERICSKB-BEST, and compare quality. For the evaluation of the

Corpus	Explanation Quality on OBQA	on QASC
QASC-17M	0.44	0.66
GENERICSKB-BEST	**0.61**	**0.79**

Figure 5: Comparative quality of two-hop explanations (sentence chains), generated using two different corpora for two different question sets[1].

chains, we'd like to coach a straightforward BERT classifier using the QASC training data (provide gold reasoning chain for each correct answer). We use the gold chains as examples of good chains, and BERT-MCQ-generated chains for incorrect answer options as samples of bad (invalid) chains. we are able to then use the trained model to judge the chains collected earlier. The results are in Figure 5, and indicate that substantially better explanations are generated with GENERICSKB-BEST. The same result was found using the OBQA dataset. particularly, because of the eclectic nature of the QASC-17M corpus, nonsensical explanations can often occur,

Example: **What do vehicles transport? people because:**
What to mention what vehicle to use
AND Now people say it's time to move on.

compared with the GENERICSKB-BEST explanation:
Example: **What do vehicles transport? people because:**
A vehicle is transport
AND Transportation is used for moving people Here, the QASC-17M explanation is nonsensical, while as GENERICSKB is rich in stand-alone generics, the reasons produced with it are more often valid.

7.3 My Evaluation

After probing all the processes in my perspective filtering rules are a touch ambiguous, it could are more clear to produce better ends up in terms of providing semantic generics. We could see that even with filtering, some contextual generics occasionally pass through the filter.

Examples include:
- All results are confidential.
- Complications are usually infrequent.

- Democracy is four wolves and a lamb voting on what to own for lunch.

These above and plenty of other examples exhibit ellipsis, vagueness, and metaphor, complicating their interpretation. Ideally, the scoring model would then score these low, but this could not always happen: recognizing contextuality often requires world knowledge. For example, consider distinguishing the great, standalone generic **Murder is illegal** from the contextual one **Parking is illegal**. Even to judge the extent of this, two annotators independently annotated 100 random (GENERICSKB) sentences from GENERICSKBBEST as to whether or not they represented useful, general truths and found 85% (averaged) met this criterion but the opposite 15% failed their criteria (but we may suggest that such problems might are relatively uncommon.)

8 Conclusion

The increasing need for general knowledge resources to improve the systems in NLP research area and a general resource for linguistics, we've created GENERICSKB, the primary large-scale resource of present generic statements, also as an augmented subset GENERICSKB-BEST, including important metadata about each statement While GENERICSKB isn't a replacement for a Web-scale corpus, we've got shown it can assist in both question-answering and explanation construction for 2 existing datasets. These positive samples of utility suggest that GENERICSKB has potential as an oversized, new resource of knowledge.

References

[1] Sumithra Bhakthavatsalam, Chloe Anastasiades, and P. Clark. Genericskb: A knowledge base of generic statements. *ArXiv*, abs/2005.00660, 2020.

[2] G. Carlson. On the semantic composition of english generic sentences. 1989.

[3] Andrei Cimpian and E. Markman. The generic/nongeneric distinction influences how children interpret new information about social others. *Child development*, 82 2:471–92, 2011.

[4] Peter Clark, Isaac Cowhey, Oren Etzioni, Tushar Khot, A. Sabharwal, Carissa Schoenick, and Oyvind Tafjord. Think you have solved question answering? try arc, the ai2 reasoning challenge. *ArXiv*, abs/1803.05457, 2018.

[5] Peter Clark, Oren Etzioni, Tushar Khot, A. Sabharwal, Oyvind Tafjord, Peter D. Turney, and Daniel Khashabi. Combining retrieval, statistics, and inference to answer elementary science questions. In *AAAI*, 2016.

[6] B. Hale and C. Wright. A Companion to the Philosophy of Language, chapter 18. 1997.

[7] Tushar Khot, Peter Clark, Michal Guerquin, P. Jansen, and A. Sabharwal. Qasc: A dataset for question answering via sentence composition. In *AAAI*, 2020.

[8] Karen T. Kohl, D. Jones, and R. Berwick. Wordnet. an electronic lexical database. edited by christiane fellbaum, with a preface by. 2003.

[9] MIT Media Lab. ConceptNet an open, multilingual knowledge graph, 1999.

[10] Todor Mihaylov, Peter Clark, Tushar Khot, and A. Sabharwal. Can a suit of armor conduct electricity? a new dataset for open book question answering. In *EMNLP*, 2018.

[11] George A. Miller. WordNet a lexical database for english, 1995.

[12] F. J. Pelletier and N. Asher. Generics and defaults. In *Handbook of Logic and Language*, 1997.

[13] Robyn Speer, J. Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge. *ArXiv*, abs/1612.03975, 2017.

[14] . Gregory n. carlson and francis jeffry pelletier (eds.), the generic book, chicago: The university of chicago press, 1995., x+463pp. 1998.