

TOWARDS AN IMPROVED LEXICON OF RELATIONAL NOUNS

ELIZABETH LEE

Department of Linguistics, University of Rochester

Abstract

The ability to detect relational nouns has shown importance in NLP tasks, particularly those that involve information extraction. For this purpose, Newell and Cheung (2018) contributed what is considered the first lexicon dedicated exclusively to relational nouns in English. In this paper, we discuss areas of improvement regarding this lexicon and propose possible steps towards an improved database for recognizing and extracting relation information from such nouns. Specific improvements that we hope to motivate are: a more encompassing list of

n

c

The ability for a system to automatically discern relational nouns would be useful for tasks involving information extraction and question answering. This goal appears to be the motivation for Newell and Cheung (2018) to create their lexicon of relational nouns. However, the decisions they made when building their lexicon raise some questions regarding the definition of a relational noun and how such a resource could actually prove useful in NLP applications.

1.1 Relational Nouns

The term *relational noun* refers to a noun that is denoted in terms of its relationship with at least one other entity. Relations expressed by relational nouns can be familial (e.g. *son*, *grandmother*), social (e.g. *friend*, *boss*), and a relative part (e.g. *edge*, *back*). The usage of relational nouns is demonstrated in sentences (1a) and (1b) below.

- (1) a. The *friendship* between Sora and Riku has lasted over ten years.
 b. Riku is Sora's *friend*.

The italicized noun in (1a) is an example of a noun which shows that there exists a relationship between two or more entities. The similar noun in (1b) denotes an entity while implying the existence of another.

As mentioned earlier, it is difficult to discern relational nouns by syntax alone as they have similar syntactic behavior to other nouns. There is one type of structure that is well-known to suggest that a noun is relational: a possessive construction of the type in which *mother* is written in the following examples. The phrase in (2a) is an example of a *prenominal possessive*. In other words, the possessor is located before the possessee in the possessive phrase and is also suffixed with 's. Example (2b) is conversely a *postnominal possessive*—more specifically a *postnominal genitive of construction*.

- (2) a. Shanna's mother
 b. The mother of Shanna

While relational nouns may often appear in these constructions, it is obvious that nouns that generally would not be considered relational often exist in identical constructions, and also that phrases of structures *x's y* and *the y of x* may not carry the same meaning.

- (3) a. Michaela's cat
 b. #The cat of Michaela

While *cat* alone may not necessarily be considered relational, in the context of (3a-b), there is a type of relationship being expressed. These constructions may also contribute to the thought that many nouns have the potential to have relational meaning attached. It has been suggested by Vikner and Jensen (2002) that non-

As far as we are aware, however, there is more to the meaning behind a relational noun that must be known in order to extract relevant and valuable information. This has largely to do with the exact argument structure of a relational noun.

1.2 How to Spot a Relational Noun

From here, our task becomes especially tricky as we come up on the task of explaining what exactly defines a relational noun. As brought up before, humans find it difficult to agree on where to draw the line between ‘relation

arguments for each noun entry suggests that this could potentially be a way either to discern relational from non-relational nouns or to split relational nouns into different categories.

While their contribution instead relates to cognitive processing, Gentner and Kurtz's (2005)

Newell and Cheung (2018) explain that a noun is relational if it is defined by virtue of how it relates to another entity. When considering which nouns to define as relational for their lexicon, they decomposed the definition of a relational noun by establishing two criteria. The first is that all relational nouns must include one of the entities in the word. For example, nouns like *brother* and *CEO* meet this requirement while others like *disagreement* do not. The second requirement is that the noun must illustrate a relation in its meaning and cannot stand alone. This restriction allows for familial terms such as *mother* but not body parts like *heart*, as one does not necessarily think of this as related to another entity. These criteria were designed to reflect the actual definition of a relational noun and exclude nouns from the lexicon that should not be considered relational under this definition.

Using their refined criteria, they created a classifier by first having a group of annotators—three experts and 13 non-experts—label nouns as belonging to one of the three categories. Whenever there was a tie from a disagreement, the label would default to “occasionally relational.” These annotations were

in accordance with their criteria specifically that *friend* and *front* are “usually relational” while *frequency* and *friendship*, while debatably relational, are labeled as “almost never relational.”

Then we come to the word *freshman*, which was perhaps given the “occasionally relational” label either because there was strong annotator disagreement or because its status as relational may depend on its context. Compare the use of this word in the sentences “*When I was a freshman in high school, I got suspended for stealing a bathroom pass*” and “*University of Rochester freshman John Smith made campus-wide fame today by uploading a selfie of himself sitting beside the so-called ‘Quad Fox’.*” We will discuss the use of this label more in depth in Chapter 3.

Moving beyond that, as we discussed briefly in the introduction, what is the extent to which a lexicon of

3.1 Must We Redefine ‘Relational’?

3.1.2 Argument-taking Nouns and Relational Nouns

The underlying problem that motivates the creation of tools to identify such lexical items is that relational nouns, which may contain valuable information regarding relations, are difficult to automatically detect while relying only on syntactic clues, as they behave the same way syntactically as sortal nouns. What we want to know from here is why we are focusing specifically on relational nouns rather than on other argument-taking nouns as well.

The categorization of nouns in NOMLEX-PLUS (Meyers 2007a, 2007b) suggests that the fundamental difference between relational and sortal nouns is the type of arguments taken. Specifically, the ARG0 of a relational noun is itself. Some nouns that might have otherwise been considered relational are labeled differently because they take an additional argument. We consider this overlap along with the fact that the classification criteria for relational nouns, in their own way, align with the traditional definition—similar to how Newell and Cheung (2018) established theirs.

3.1.3 How to Determine What is Relational

Taking the details in this section into consideration, we describe here what we believe constitutes a noun as relational.

We consider **all argument-taking nouns** to be relational. The mindset behind this is that if a noun takes an argument, it can be argued that there exists some relationship between the predicate and its argument. Furthermore, regarding applications in NLP, the information attached to arguments of such nouns seems just as crucial as that of arguments of other nouns.

Argument-taking nouns as relational nouns can be separated by category based on their structure. We might have a category containing nouns that clearly meet Newell and Cheung's criteria—what we might call 'traditional' relational nouns—as well as categories for nominalizations of verbs and adjectives, part-whole terms, properties and attributes, and so on.

We do admit, however, that this definition is not without its complications. Mainly, we are aware of the fact that argument-taking can be “forced” upon any noun, and that any sortal noun acting as the possessee in a possessive construction can be given some relation to the possessor. With this fact, we come back around to the claim that most if not all nouns may possess some

In section 5, we explain why we do not attempt to handle polysemy given the resources we choose to employ. Hence we leave improvement on this particular aspect as a task for future work.

4 Relevant Works

There are few works aside from Newell and Cheung (2018) that have specifically dealt with the handling of relational nouns as they

5 Pilot Studies

importance values from this model in Table 4. Note that when evaluating for average precision and F1 score, we treated all “occasionally relational” labels as “usually relational” ones.

Table 3. Average precision (AP) and F1 scores of Study One in percentages.

Table 4. Top ten important features by as indicated by our random forest model in Study One. Feature importance values were rounded to the nearest thousandth.

As per evidence from the literature, relational nouns often exist in possessive and appositive constructions. Other features include occurring with prepositional phrases and definite articles. Although we see that verb constructions seem to be a common trend here, we wonder if this merely has something to do with the general frequency of verb phrases.

It should be kept in mind that these important features were determined based on labels from Newell and Cheung’s lexicon, which were influenced by their constrained definition of relational nouns. We attempt to obtain a more pertinent set of features through our procedure in the next study.

We additionally provide a few examples of newly labeled relational nouns in Table 5 as determined by the random forest classifier. Predictably, we get some part-of (*cog*), social (*combatant*), and property (*inanimateness*) terms. We do, how

Table 5. Ten random nouns labeled as “usually relational” by the random forest classifier in Study One.

In terms of the number of nouns labeled, 137 were labeled as “occasionally relational” while 697 were labeled as “usually relational.” This gives us a total of 834 new nouns with relational labels. Combined with the nouns given labels from the lexicon, we get a total of 5,831 relational nouns. We may have ended up with fewer than we started, but we are not finished here yet.

5.3 Study Two

The procedure of this study is similar to that of Study One, but different only in the labeling mechanisms. In the hopes that we could obtain new labels for nouns based on our new criteria for relational nouns, we altered the ‘gold’ labels in our dataset to meet this definition. To aid us in this task, we extracted the argument-

Table 7. Top ten important features correlating with relational nouns as indicated by our random forest model in Study Two. Values were rounded to the nearest thousandth.

In the top features of this round, we see that being a prepositional subject comes out on top, with a significantly higher score than the rest on the list (Table 7). While determiners and possessives make the list again, we have some interesting new features, like occurrence with noun negations, adjectives, numerals, and adverbs.

Table 8. Ten random nouns labeled as “usually relational” by the random forest classifier in Study Two.

We will start by taking the relational nouns as labeled in Study Two and sorting them into subcategories based on the groups we mentioned in §3.1.3 using a supervised clustering algorithm. Once those clusters are formed, we would ask human annotators to evaluate the nouns placed in each category to determine the effectiveness and accuracy of the clustering. If these nouns appear to be accurately sorted, we can say that we made a step towards improving the original lexicon by incorporating some information regarding argument structure.

5.5 Discussion

5.5.1 Grammatical Features of Relational Nouns

Here we examine commonly observed behaviors of relational nouns, partly to provide some explanation for the particular features being considered significant by our models and partly to explore ways of identifying this type of '

Third, we did not have enough time to let human annotators thoroughly evaluate the generated lists of relational nouns and their clusters before this paper was finished. Hence at the moment we cannot say for certain whether or not our own lexicon could be a reliable resource. Hopefully, the right evaluations will be carried out in the near future as we continue and expand on this project.

The relabeling of nouns for Study Two was automated rather than hand-annotated. One step in the revision of these procedures in the future should be to evaluate the relabeled nouns and ensure that they actually align with our new criteria for considering nouns as relational.

A final point to note is that the dataset of nouns which we used to build our lexicon does not differentiate nouns based on word sense—each entry only corresponds to a different lemma. Because of this, we are unable to move forward with our goal of labeling each sense of a noun as opposed to each lemma while relying on our particular dataset.

6 Conclusion

The ability to detect relational nouns has shown importance in NLP tasks, particularly those that involve information extraction. Following the lead of Newell and Cheung (2018), we look towards developing an improved lexicon of relational nouns by considering the shortcomings of their 2018 lexicon, using this knowledge when discussing methods for building a new one, and presenting pilot data as an example of how to take the next step in future work. Specific improvements that we hope to achieve include a more encompassing list of relational nouns based on a broader definition of the term, more informative entries based on the argument structures of each noun, and the ability to recognize and handle polysemy of nouns that yield both relational and non-relational interpretations. We hope to continue on this work in the future as well as to motivate others to take these points into consideration when moving forward with tools for relation extraction, and that whatever comes of this effort will be useful for various natural language processing tasks.

6.1 FuturF6

Acknowledgments

I would first like to thank Dr. Scott Grimm, my advisor for this project and a fantastic professor. Thank you for being the one to introduce me to the wonderful world of semantics. I also want to thank the Linguistics department faculty for their helpful comments during my presentation for this project.

Next, I want to thank to classmates Hannah An, Venkat Govindarajan, and Siddharth Vashishtha. I have been incredibly lucky to be part of a cohort with some of the coolest people I have met, and I could not have made it without your kindness and support (and baked goods). I would also like to thank computer science graduate students Aeshaan Wahlang and Rik Bose for their technical and writing advice.

Finally, I want to thank my family – Peter, Patricia, and Jamie Lee – for supporting me through these past two years. I love you so much.

References

- Asudeh, Ash. 2005. Relational nouns, pronouns, and resumption. *Linguistics and Philosophy* 28.4: 375–446.
- Barker, Christian. 1995. *Possessive Descriptions*. CSLI Publications, Stanford, CA.
- Barker, Christian. 2011. Possessives and relational nouns. in *Semantics: An International Handbook of Natural Language Meaning*, 1108-1129. De Gruyter.
- Davies, Mark. 2008. *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. url: <https://corpus.byu.edu/coca/>.
- de Bruin, Jos and Remko Scha. 1988. The Interpretation of Relational Nouns. *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, 25–32. Association for Computational Linguistics.
- Fellbaum, Christiane.

- Newell, Edward and Jackie Chi Kit Cheung. 2018. Constructing a Lexicon of Relational Nouns. in Nicoletta Calzolari (ed.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA). isbn: 979-10-95546-00-9.
- Pal, H. and Mausam. 2016. Demonyms and compound relational nouns in nominal open IE. *Proceedings of AKBC*, 35–39.
- Pedigrosa, F. et al. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
- Staroverov, Peter. 2007. Relational nouns and reciprocal plurality. *Semantics and Linguistic Theory* 17, 300–316.
- Vikner, Carl and Per Anker Jensen. 2002. A semantic analysis of the English genitive. Interaction of lexical and formal semantics. *Studia Linguistica* 56.2, 191–226.