

A sentiment-annotated dataset of English causal connectives

Marta Andersson*, Murathan Kurfali†, Robert Östling†

*English Language Department, Stockholm University, Stockholm, Sweden

†Linguistics Department, Stockholm University, Stockholm, Sweden

marta.andersson@english.su.se

{murathan.kurfali, robert}@ling.su.se

Abstract

This paper investigates the semantic prosody of three causal connectives: *due to*, *owing to* and *because of* in seven varieties of the English language. While research in the domain of English causality exists, we are not aware of studies that would cover the domain of causal connectives in English. Our claim is that connectives such as *because of* link two arguments, (at least) one of which will include a phrase that contributes to the interpretation of the relation as positive or negative, and hence define the prosody of the connective used. As our results demonstrate, the majority of the prosodies identified are negative for all three connectives; the proportions are stable across the varieties of English studied, and contrary to our expectations, we find no significant differences between the functions of the connectives and discourse preferences. Further, we investigate whether automatizing the sentiment annotation procedure via a simple language-model based classifier is possible. The initial results highlights the complexity of the task and the need for complicated systems, probably aided with other related datasets to achieve reasonable performance.

1 Introduction and background

Examination and extraction of sentiment (*sentiment analysis*, SA) from text traditionally rely on the coarse distinction between positive and negative polarity. The mainstay of the current methodologies is the idea that a text contains a single sentiment about one single topic, as in (1) below (Benamara et al., 2017; Mohammad, 2016). However, it is now generally acknowledged that sentiment within one topic should be considered at several levels of granularity, as in (2):

- (1) *Out of Africa* is a fantastic movie.
- (2) *Out of Africa* is a fantastic movie but a boring book.

This simple concessive construction in (2), which conveys both a positive and a negative sentiment about the same topic, demonstrates that the same word can be assigned different polarity labels dependent on the domain and context (cf. *terrible thriller* and *terrible weather* (Zitoune et al., 2016)), or on the nature of other discourse phenomena present in its close textual environment (cf. *You must read this paper* and *I believe you must read this paper* — in the latter case the negative polarity is reduced by the hedging effect of the verb *believe*). Observations of this kind have spawned interest among SA researchers to tackle phenomena such as semantic gradation or context/domain influence on evaluation; however, an approach to sentiment detection which would allow for a theoretical characterization of text phenomena from the linguistic point of view, is much needed in the field (Mohammad, 2016).

One domain that is both challenging and useful for sentiment detection is causality — notoriously difficult to analyze (in spite of being regarded as one of so-called “semantic primitives” (Wierzbicka, 1998)), as it operates both at the level of factual events in the real world, and at the level of “meta-causality” — i.e. the speaker’s reasoning and exchange between interlocutors (e.g. conclusions and

This work is licensed under a Creative Commons Attribution 4.0 International License. Licence details: <http://creativecommons.org/licenses/by/4.0/>.

speech acts; see (Sweetser, 1990) and the literature on computational methods to detect causality, e.g. (Kang et al., 2017)). One example of a relation that conveys causality at the level of the speaker's reasoning is (3) below:

- (3) the publications themselves are a kind of poetic transformation *due to* the finely crafted nature of them (Jamaican Eng.)

Both the causal argument (2nd clause) and the result argument (1st clause) convey the speaker's opinion, i.e. not a result based on a cause in the physical world. While such relations are very common in discourse, producing this type of deep causal explanation is beyond the capability of the existing systems (Kang et al., 2017). Attempts have been made in the field to capture causal explanations as, first of all, temporal sentiment analysis to predict causal relations (Kang et al., 2017; Preethi et al., 2015), and sentiment analysis and causal rule detection (Dehkharghani et al., 2014; Siganos et al., 2014). Some studies also utilize discourse connectives as predictors of discourse relation types, for instance, Wang et al. (2012) and Mukherjee and Bhattacharyya (2012), and indicate that discourse particles such as *but*, *since*, or *although* can be used to improve sentiment classification accuracy. Importantly, several linguistic studies have proved that discourse connectives prime or at least support the intended discourse interpretations as real-world consequences or meta-linguistic effects such as conclusions, opinions or speech acts (Andersson, 2019; Kamalski et al., 2008; Scholman and Demberg, 2017).

Since causality is basic to both human cognition and discourse coherence, our paper intends to further explore the nature of causal explanations and the role of discourse connectives in relation disambiguation by focusing on three causal connectives in English — *due to*, *because of* and *owing to* — from the point of view of their semantic prosodies. Semantic prosody has been described at several levels of abstraction (e.g., affective meanings of a given node with its typical collocates) (Sinclair, 1998; Stubbs, 2001), the discussion of which would be beyond the scope of this paper. The view adopted in the following is that semantic prosody is a feature of the node word that dictates the general environment which constrains the preferential choices of this word. As a result, it can affect wider stretches of text and so the words often tend to co-occur with either 'negative' ('bad', 'unpleasant') or 'positive' ('good', 'pleasant') collocates (Partington (2004), see also Xiao and McEnery (2006)). A related idea is colligation, i.e. the relationship between a node and grammatical categories. This feature is, however, excluded from our analysis based on the observation that all the investigated connectives follow the standard syntactic pattern: connective + NP in their target senses (Xiao and McEnery, 2006), for instance:

- (4) Afghanistan and Iraq today is a bloody mess **because of** the Westminster's style of diplomacy. (British Eng.)

In order to detect the sentiment of the analyzed connectives, our first step was to investigate their general collocational behavior. To this end, we started our study by consulting the Oxford English Dictionary (OED).¹ While the connectives all have synonymous surface senses, their meanings and functions in natural discourse can be perceived as either negative, positive or neutral. According to the OED, the etymologies of the expressions that the connectives stem from can be assessed as quite negative for both *due to* and *owing to* (i.e. "indebted or beholden" for *owe* and "debt or obligation" for *due*), and neutral for *because* ("for the reason that"). However, the target senses we are interested in (i.e. the connective as a compound preposition followed by an NP), are all described as relatively neutral: while *because of* means "by reason of; on account of"; *due to* is defined as: "as a result of, on account of, because of"; and the target meaning of the connective *owing to* has been described as: "in consequence of, on account of, because of". We can therefore assume that our connectives are intrinsically neutral. This assumption is important in a study of semantic prosody, since words that have clear negative or semantic prosodies are hard to investigate given their context-neutralizing function (e.g. the verb *alleviate* remains positive even if accompanied by a clearly negative phrase, such as *suffering* (Lin and Chung, 2016)).

The conclusion to draw from this part of the investigation is twofold — while the analyzed connectives indeed seem to have quite neutral senses, given their complex etymologies and quite specific semantics

¹<https://www.oed.com/>

(as opposed to that of multi-functional connectives such as *so*), it can be assumed that they will show tendencies to occur with either favorable or unfavorable events. Based on the previous studies within the domain of causality that have indicated that the verb *cause* has a negative semantic prosody (e.g. Stubbs (1995)), which has been demonstrated to prime experimental subjects to think about the same event as negative if preceded by this verb (Hauser and Schwarz, 2016), it can be assumed that at least some of the connectives in question will exhibit negative collocational behaviour. Needless to say, language users may quite flexibly establish the linguistic environment of any phrase (connectives included), based on their communicative and rhetorical purposes (cf. Lin and Chung (2016), on the prosody of the word *challenge*).

In order to establish the type of semantic prosody related to the analyzed connectives, we investigated their immediate linguistic environments and assessed whether the node item occurs with favourable or unfavourable meanings. While the underlying idea was that a specific word/phrase would be found in the context as governing the positive or negative flavour of the sentence, we analyzed at least a whole clause preceding and following the connective.

Apart from the practical use of our computational method for automatic annotation of the sentiment of discourse connective arguments, our investigation gives us a more detailed understanding of the differences between these near-synonymous connectives in English. Our aim has also been to use such differences as a proxy to investigate pragmatics-level language change, by making quantitative comparisons between different English varieties. The same methods could be applied to diachronic studies, given a corpus with sufficiently comparative text over a suitable time span. This goes beyond the scope of the current project.

2 Research Questions

While our main contribution in this paper is the annotated resource as such, the original motivation for performing the annotation work is to answer a number of questions on semantic prosody of causal connectives.

1. Do the different connectives in our study (*because of*, *due to* and *owing to*) display differences in their semantic prosodies? Our working hypothesis is that their discourse functions/preferences will differ and so, based on these differences, the connective can serve as a predictor of the sentiment.
2. Do the semantic prosodies of causal connectives differ between varieties of English? In particular, certain corpus observations prior to this study suggested that the negative connotations of *due to* may be less pronounced in some Asian varieties. We therefore intend to check the extent of usage flexibility, which may affect the predictive power of the connective.
3. Are the connotations of each connective more closely connected with one argument than the other? This would show us a non-even distribution between negative-cause/positive-effect and positive-effect/negative-cause uses, and allow a more fine-grained way of studying the use and evolution of causal connectives.

3 Corpus

3.1 Collection of dataset and Method

Randomly collected samples of the Global Web-Based English (GloWbE; (Davies, 2013)) corpus were annotated using PDTB Annotator (Lee et al., 2016). GloWbE contains about 1.9 billion words of text from twenty different countries. While the texts in the corpus consists of informal blogs (about 60% of corpus) and other web-based material, such as newspapers, magazines, and company websites, our study excludes the blogs, as they were available only in two of the analyzed varieties. It should also be noted that this usage of discourse connective is an extension to some approaches to discourse which require connectives to link two sentences (e.g., PDTB). In contrast, the connectives under study often connect a sentence to an NP.

As pre-processing, for each connective, the context of 250 words from left and right was extracted automatically. A random subset of word tokens has been erased from the GlobWbE corpus due to copyright reasons, so we only consider instances where the full context is present. We also discard all instances where a connective occurs more than once.

The following section describes the details of the annotation method and the decisions made in the process of corpus coding.²

3.2 Annotation Guidelines

Our annotation task can be regarded as sentence-level annotations of causal sentiment similar to Rosenthal et al. (2015) and Mohammad et al. (2016), who framed the task as “is this sentence positive, negative or neutral?” However, as mentioned, our idea is that the choice of discourse connective to signal a specific causal event type is governed by the presence of a discourse entity(ies) (underlined in the following) that can be coded as positive or negative based on its semantic nature. Such an entity may be but does not have to be an entire clause/sentence. For instance:

- (5) When the examiners award you the degree at the end of your viva and you emerge out into the street near to tears **because of** tension/tiredness/relief (...) (British Eng.)

The two annotators of the AmE samples of the connective *because of* consistently (see Section 4.1 below) agreed that the underlined is a minimal text span both needed and sufficient to be identified as an antecedent/postcedent of the connective (in this case – both negative). This is consistent with the PDTB manual on the connective annotation, according to which “only as many clauses and/or sentences should be included in an argument selection as are minimally required and sufficient for the interpretation of the relation” (Prasad et al., 2007). Many naturally produced examples are, however, less straightforward and involve conceptually more complex arguments. Consider:

- (6) It’s very good to accomplish this **due to** the fact you’ll remain present and understand what your choices are (Nigerian Eng.)

Why such relations are more difficult to analyze is because it is the entire event (and consequently, the whole clause) that should be treated as an antecedent or/and postcedent of the connective. This differs from the straightforward context of (5) with NPs as ante-/postcedents. Interestingly, (5) and (6) demonstrate that the analyzed connectives can signal both clearly negative and also clearly positive events. The question therefore arises on how to treat relations where each argument evokes different event type/includes different antecedent type. Consider:

- (7) He’s cognizant to make sure the proper people are credited this time **because of** what he went through last time (Malaysian Eng)

While the argument following the connective clearly implies a negative experience, the preceding argument conveys a positive outcome. This presents a methodological problem of how to annotate the nature of the whole event — in some approaches, the argument of an explicitly marked relation (as opposed to juxtaposed sentences) are maximally interpretable only in the context of another argument (Blakemore and Carston, 2005). This means that both arguments contribute to the interpretation of the relation. While this is a plausible hypothesis, which would explain the presence of the intrinsically neutral connective *because of* in this context, we believe that such claims have to be supported by experimental evidence (e.g. acceptability judgements). We have therefore decided to annotate the arguments separately and assess the prosody based on one argument being either positive or negative. However, most relations include either two arguments of the same type or one that appears neutral:

- (8) My son’s cat became diabetic and was told by his vet that they are seeing more cases **due to** cats being fed dry food. (British Eng.)

²The corpus is available at <https://github.com/MurathanKurfali/sentimental-causal-connectives>

	because of	due to	owing to	Sum
American	99	97	97	293
Australian	98	–	–	98
British	99	89	97	285
Indian	99	94	–	193
Jamaican	–	96	60	156
Malaysian	95	97	–	192
Nigerian	–	91	97	188
Sum	490	564	351	1405

Table 1: Number of annotated relations per connective and per English variety.

While feeding animals dry food is not negative *per se*, in the context of its resulting in diabetes, the event becomes negatively tinted. This example is thus suggestive as to interpreting discourse entities at face value.

Nevertheless, as mentioned in Section 1 above, detecting causal explanations from texts is a complex task. Particularly difficult to analyze are relations that involve the speaker’s emotional state, sarcasm and ridicule, rhetorical questions etc., see e.g. Mohammad (2016). Yet, our corpus material seems to consist mostly of relations that can be regarded as “neutral reporting of valenced information” (Mohammad, 2016), as (8) above. This means that it may be hard to judge whether the events should be regarded as neutral reporting or rather a negative emotional state, since the speaker does not provide any indication of her own emotional state. In (8), is the speaker just reporting on the course of events or upset about the situation? Since there are no overt signals of the speaker’s emotional state, it seems that even more sophisticated methods of relation extraction would have difficulties assigning sentiment to (8). As a solution to this problem, we followed (Mohammad, 2016) simple sentiment questionnaire and annotated such relations as “the speaker is neither using positive language nor using negative language”.

One final remark is that the rather neutral nature of the causal relations in our corpus samples may be related to both the nature of the corpus, which consists of web-based materials, such as: company websites, newspapers, magazines etc., which are not the primary site of the speaker’s subjective opinions, and also to the very nature of unambiguous English connectives, which have been demonstrated to be used for very specific purposes (Andersson, 2019). Based on the semantics of the connectives analysed here, we assume that they are mostly used to signal factual event types. Yet, more research should be carried out to confirm this hypothesis.

3.3 Annotation Statistics

Each connective was annotated in at least four varieties of English to capture any possible differences in terms of the contexts they occur in (see Table 1). Our aim has been to code 100 relations for each connective, although in most cases a small number of examples was excluded because there was insufficient context to perform annotations. In one case, “owing to” in JmE, the number of annotations is significantly lower, since only 60 examples were found in the corpus.

Note that in the following, we will refer to the effect as *argument 1* and cause as *argument 2* (and order them as such when labels are paired). This choice reflects the default order of events in English, but inversion of the clauses is frequent in actual use. We then adhere to our convention by annotating *argument 2* as preceding *argument 1*.

4 Analysis

4.1 Inter-Annotator Agreement

To test the annotation guidelines, *because of* instances in the US corpus is annotated blindly by two annotators additional to the main annotator, out of whom the second annotator is more experienced than the third one annotating more files in total. We have selected *because of* for inter-annotator agreement (IAA) as our initial hypothesis is that it is more likely to occur in both negative and positive contexts

due to										
Variety	neg/neg	neg/neu	neg/pos	neu/neg	neu/neu	neu/pos	pos/neg	pos/neu	pos/pos	TOTAL
American	44	11	3	4	17	1	2	6	9	97
British	41	15	3	8	7	1	2	8	4	89
Indian	52	16	1	6	5	3	1	4	6	94
Malaysian	42	19	2	13	9	0	2	7	3	97
Nigerian	44	11	0	9	11	5	1	5	5	91
Jamaican	42	15	2	7	11	2	2	9	6	96
because of										
Variety	neg/neg	neg/neu	neg/pos	neu/neg	neu/neu	neu/pos	pos/neg	pos/neu	pos/pos	TOTAL
Indian	25	20	0	13	7	4	4	17	9	99
British	38	17	3	12	9	3	3	10	4	99
Malaysian	21	15	3	7	15	4	3	15	12	95
Australian	21	19	0	16	24	7	1	6	4	98
American	33	8	5	8	14	5	5	7	14	99
owing to										
Variety	neg/neg	neg/neu	neg/pos	neu/neg	neu/neu	neu/pos	pos/neg	pos/neu	pos/pos	TOTAL
American	29	17	3	4	19	3	1	5	16	97
Jamaican	16	11	0	5	11	2	1	6	8	60
British	39	19	3	6	14	2	1	6	7	97
Nigerian	46	11	2	4	8	6	2	4	14	97

Table 2: Distribution of labels in our annotated data. For instance, the category neg/neu means that the label of argument 1 (effect) is *negative*, while argument 2 (cause) is *neutral*.

Annotation	1-2	1-3	2-3
1st argument (effect)	0.816	0.569	0.609
2nd argument (cause)	0.648	0.348	0.345
Combined	0.788	0.544	0.537

Table 3: Inter-annotator agreement for the sentiment annotations of each argument individually, as well as of the whole pair counted as one unit between all annotators.

than the others, hence more challenging to annotate. We calculate the IAA on each argument separately as well as on the whole relation using linearly weighted Cohen’s Kappa (McHugh, 2012) between each annotator pair.

Table 3 shows that for each case, IAA results between the most experienced annotators are ≥ 0.6 which is regarded as substantial (Cohen, 1960).

The lower level of inter-annotator agreement between the third annotator and the others indicates the complexity of the analyzed relations and highlights a need for a training period. Yet, the IAA scores alone fail to provide any insight about the nature of agreement between annotators, hence we also analyze how the annotators disagree with each other. As can be seen in Figure 1a to Figure 1c, the disagreements are overwhelmingly between *Neutral* and one of the polarities. There are only few instances where annotators assign opposite polarities.

4.2 Differences Across Varieties and Connectives

Our preliminary observations indicated that *due to* may have less negative connotations in Asian Englishes, represented in our sample by Indian and Malaysian English. We are therefore interested in testing this hypothesis, as well as the more general question of whether there are differences in general between the English varieties with respect to the different connectives.

For the specific question of *due to*, we define a *negative context* as any relation with at least one negative and no positive argument, i.e. negative/negative, negative/neutral or neutral/negative. We also define *non-negative contexts*, consisting of all relations with no negative argument. We compare two groups: *Asian* (Indian and Malaysian) English, and the *major* (American and British) English varieties. Using these definitions, we see that contrary to our hypothesis, the major varieties have a somewhat *smaller* (70%, 123 of 176) proportion of negative uses than the Asian varieties (80%, 148 of 185). This difference is statistically significant, but only marginally so (χ^2 test, $p = 0.04$).

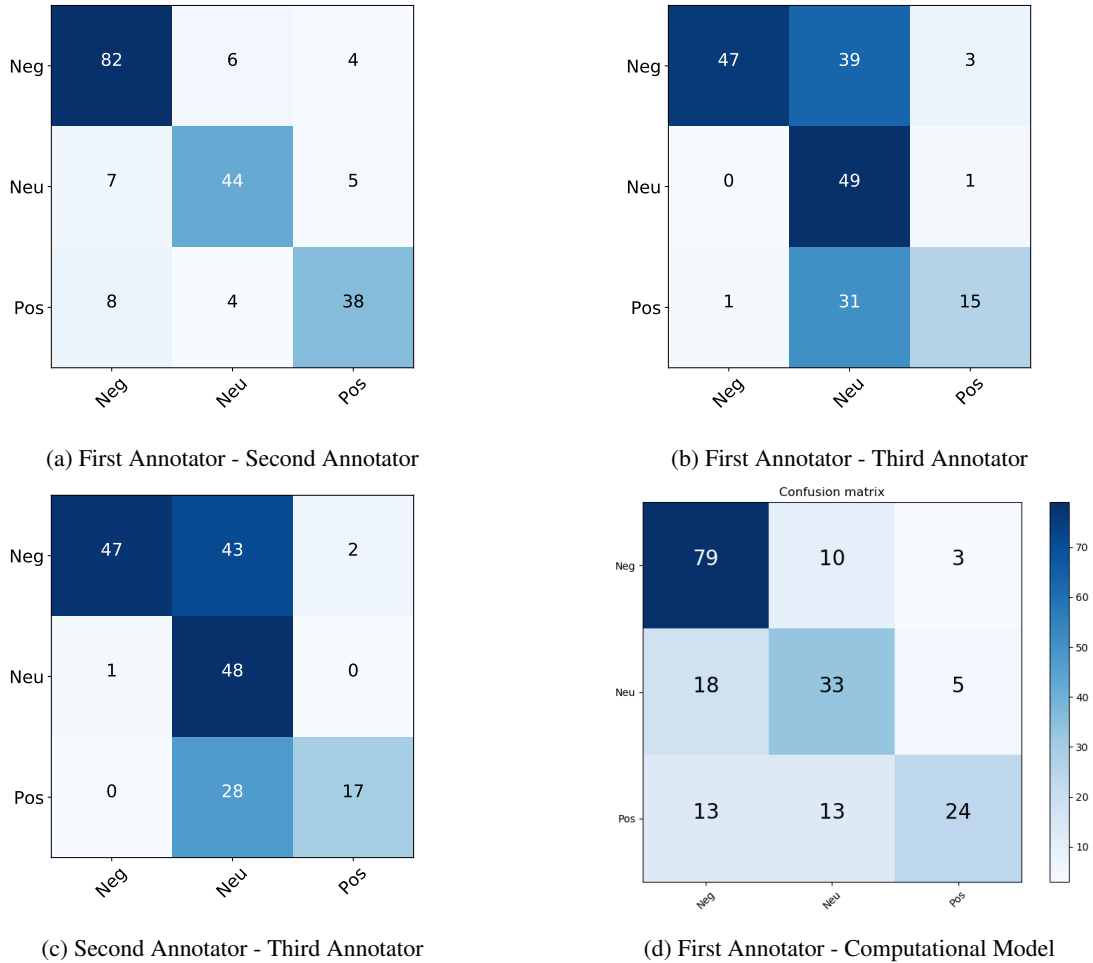


Figure 1: Confusion matrices between each annotator pair as well as between the first annotator and the computational model.

For the more general question, we compare all varieties that have been annotated for a given connective, with respect to three categories: *positive* (positive/positive, positive/neutral, neutral/positive), *neutral* (neutral/neutral), and *negative* (negative/negative, negative/neutral, neutral/negative). We find that there is no significant difference (χ^2 test, $p = 0.26$) for *due to*, whereas both *because of* and *owing to* display significant differences between varieties (both have $p < 0.001$).

Although some differences are statistically significant, they are relatively small. Below, we summarize the annotation statistics (available in their entirety in Table 2):

- *due to*: large majority of negative (64–80%) uses, some neutral (5–18%) and positive (11–18%) uses.
- *because of*: majority of negative (48–72%), some neutral (7–25%) and many positive (18–35%) uses.
- *owing to*: majority of negative (54–69%), some neutral (9–20%) and many positive (16–27%) uses.

The ranges within parentheses indicate the minimum and maximum proportion of instances for that particular category, across the varieties for which we have annotations.

4.3 Differences Between Arguments

As shown in Table 4, cases where the arguments are of opposite polarity (negative/positive and positive/negative), are quite rare. Across all 1405 annotated relations, only 30 (2.1%) are annotated as

	neg/pos	pos/neg
due to	11	10
because of	11	13
owing to	8	5
Sum	30	28

Table 4: Number of relations with arguments of opposite polarity.

negative/positive, and 28 (2.0%) as positive/negative. These have been left out in the comparisons in Section 4.2, because it is not straightforward to classify the whole relation as either positive or negative. However, these rare cases provide an opportunity to test our third research question: are the connotations of each connective more closely connected with one argument than the other?

We find that both opposition pairs (negative/positive and positive/negative) are about equally well-represented for each of the connectives, speaking against the hypothesis that predicts a much stronger connection of the connective to one argument over the other. However, more corpus research would be needed to verify/discard this hypothesis more reliably. See also the discussion on example 7, which has been annotated as positive/negative.

4.4 Computational Model

Along with the manual annotations, we also explore the automatic means to annotate sentiments of the arguments. To this end, we model the problem as a sentence classification task and fine-tune a pre-trained language model, BERT (Devlin et al., 2018),³ which has become the standard procedure in NLP. As the task here is to predict the sentiment of an argument, we only pass the annotated text spans to BERT. The computational model is implemented using Huggingface’s Transformers library.⁴ As fine-tuning procedure is known to suffer from high variance, the model is run 5 times and we report test results of the run with the median development performance.

To test the performance of the classifier, we allocated the *US English because of* annotations as the test data and train the classifier on the remaining annotations.⁵ In order to understand how well the classifier performs, we report the IAA scores between the classifier and each annotator in Table 5.

Annotation	Ann1	Ann2	Ann3
1st argument (effect)	0.504	0.529	0.300
2nd argument (cause)	0.507	0.376	0.235

Table 5: Inter-annotator agreement (Cohen’s Kappa) between the computational model and the annotators.

The results show that the classifier cannot match the human performance. When the divergences between the predicted labels and first annotator’s coding (according to which the classifier is trained) are examined, we see that the nature of these disagreements is different from that of human annotators (see Section 4.1), as the model is less likely to label the arguments as *positive*. Yet, the total number of disagreements between opposite polarities is almost the same (compared Figure 1a and 1d), which suggests that the computational model does not make extremely precise predictions.

In almost all cases, however, the model’s agreement with the experienced annotators are moderate, with $\text{Kappa} \geq 0.4$, which we find promising given the limited number of the annotations available for training.

³We use bert-large-cased model.

⁴<https://github.com/huggingface/transformers>

⁵which consists of 2612 instances in total

5 Conclusion

The main finding of our study is that the three causal connectives analyzed, *due to*, *because of* and *owing to*, are all associated with predominantly negative semantic prosodies. Other than this general trend, the distributions of usages do not significantly differ between the connectives, between the analyzed regional varieties of English, or between the arguments (cause and effect).

We investigated the possibility of automating sentiment annotations for this project in Section 4.4 and our results indicate that the computational model achieves moderate agreement with the annotators despite the limited training data, suggesting that the procedure can benefit more annotations and have potential to be automatized to a good extent after a certain amount of manual annotation. The topics for future work are to involve more human annotators and develop more precise guidelines based on their observations and, also, investigate whether other, larger, sentiment-related resources can be exploited for this task.

References

- Marta Andersson. 2019. Subjectivity of english connectives. *Empirical Studies of the Construction of Discourse*, 305:299.
- Farah Benamara, Maite Taboada, and Yannick Mathieu. 2017. Evaluative language beyond bags of words: Linguistic insights and computational applications. *Computational Linguistics*, 43(1):201–264.
- Diane Blakemore and Robyn Carston. 2005. The pragmatics of sentential coordination with and. *Lingua*, 115(4):569–589.
- Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.
- Mark Davies. 2013. Corpus of global web-based english: 1.9 billion words from speakers in 20 countries (glowbe). Available online at corpus.byu.edu/glowbe/. Retrieved December, 15:2015.
- Rahim Dehkharghani, Hanefi Mercan, Arsalan Javeed, and Yucel Saygin. 2014. Sentimental causal rule discovery from twitter. *Expert Systems with Applications*, 41(10):4950–4958.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- David J Hauser and Norbert Schwarz. 2016. Semantic prosody and judgment. *Journal of Experimental Psychology: General*, 145(7):882.
- Judith Kamalski, Leo Lentz, Ted Sanders, and Rolf A Zwaan. 2008. The forewarning effect of coherence markers in persuasive discourse: Evidence from persuasion and processing. *Discourse Processes*, 45(6):545–579.
- Dongyeop Kang, Varun Gangal, Ang Lu, Zheng Chen, and Eduard Hovy. 2017. Detecting and explaining causes from text for a time series event. *arXiv preprint arXiv:1707.08852*.
- Alan Lee, Rashmi Prasad, Bonnie Webber, and Aravind K Joshi. 2016. Annotating discourse relations with the pdtb annotator. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pages 121–125.
- Yen-Yu Lin and Siaw-Fong Chung. 2016. A corpus-based study on the semantic prosody of challenge. *Taiwan Journal of TESOL*, 13(2):99–146.
- Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, 22(3):276–282.
- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41.
- Saif Mohammad. 2016. A practical guide to sentiment annotation: Challenges and solutions. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 174–179.

- Subhabrata Mukherjee and Pushpak Bhattacharyya. 2012. Sentiment analysis in twitter with lightweight discourse analysis. In *Proceedings of COLING 2012*, pages 1847–1864.
- Alan Partington. 2004. ”utterly content in each other’s company” semantic prosody and semantic preference. *International journal of corpus linguistics*, 9(1).
- Rashmi Prasad, Eleni Miltsakaki, Nikhil Dinesh, Alan Lee, Aravind Joshi, Livio Robaldo, and Bonnie L Webber. 2007. The penn discourse treebank 2.0 annotation manual.
- Peter G Preethi, Vilma Uma, et al. 2015. Temporal sentiment analysis and causal rules extraction from tweets for event prediction. *Procedia computer science*, 48:84–89.
- Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif Mohammad, Alan Ritter, and Veselin Stoyanov. 2015. Semeval-2015 task 10: Sentiment analysis in twitter. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pages 451–463.
- Merel Scholman and Vera Demberg. 2017. Crowdsourcing discourse interpretations: On the influence of context and the reliability of a connective insertion task. In *Proceedings of the 11th Linguistic Annotation Workshop*, pages 24–33.
- Antonios Siganos, Evangelos Vagenas-Nanos, and Patrick Verwijmeren. 2014. Facebook’s daily sentiment and international stock markets. *Journal of Economic Behavior & Organization*, 107:730–743.
- John Sinclair. 1998. The lexical item. *Amsterdam Studies In The Theory And History Of Linguistic Science Series 4*, pages 1–24.
- Michael Stubbs. 1995. Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of language*, 2(1):23–55.
- Michael Stubbs. 2001. *Words and phrases: Corpus studies of lexical semantics*. Blackwell Publishers Oxford.
- Eve Sweetser. 1990. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*, volume 54. Cambridge University Press.
- Fei Wang, Yunfang Wu, and Likun Qiu. 2012. Exploiting discourse relations for sentiment analysis. In *Proceedings of COLING 2012: Posters*, pages 1311–1320.
- Anna Wierzbicka. 1998. Anchoring linguistic typology in universal semantic primes.
- Richard Xiao and Tony McEnery. 2006. Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied linguistics*, 27(1):103–129.
- Farah Benamara Zitoune, Nicholas Asher, Yvette Yannick Mathieu, Vladimir Popescu, and Baptiste Chardon. 2016. Evaluation in discourse: A corpus-based study.