

# Automatic Animacy Classification for Romanian Nouns

Maria Tepei, Jelke Bloem

Institute for Logic, Language and Computation, University of Amsterdam  
maria.tepei@student.uva.nl, j.bloem@uva.nl

## Abstract

We introduce the first Romanian animacy classifier, specifically a type-based binary classifier of Romanian nouns into the classes human/non-human, using pre-trained word embeddings and animacy information derived from Romanian WordNet. By obtaining a seed set of labeled nouns and their embeddings, we are able to train classifiers that generalize to unseen nouns. We compare three different architectures and observe good performance on classifying word types. In addition, we manually annotate a small corpus for animacy to perform a token-based evaluation of Romanian animacy classification in a naturalistic setting, which reveals limitations of the type-based classification approach.

## 1. Introduction

Animacy is a semantic property of nouns that describes the quality of the noun's referent of being alive, sentient or volitional. Reference to animate entities thus comprises the usage of proper names, profession names, institutions, and company names. In a variety of languages, animacy plays a role in determining the inflection and agreement of nouns, particularly when it comes to grammatical gender or case marking. Having comprehensive tools for classifying nouns based on this feature has important implications for modelling a language's grammar, as well as for natural language processing (NLP) tasks such as machine translation and text analysis.

Classification of nouns by animacy is a small but active area in NLP, for various languages. For instance, earlier research has shown that animacy plays an important role in the inflection of Russian nouns (Pereltsvaig and Battistella, 2006), while Sanchez-Gomez et al. (2011) show that animacy-based classification of Spanish nouns can be used to improve machine translation, a perspective that is also supported by Øvrelid (2005). Conversely, Romanian is a relatively under-resourced language when it comes to this area of research and to NLP in general. Despite being a Romance language with a diverse noun declension pattern, we are not aware of any approach to this date focusing Romanian animacy classification. One reason for this gap might be the fact that, despite this feature being an important one in Romanian grammar, animacy is not marked morphologically to the same extent as in other languages (see Pereltsvaig and Battistella, 2006 for Russian or Shah et al., 2020 for Hindi). Thus, it cannot be used as a feature for NLP tasks as easily as for languages with frequent explicit animacy marking. However, animacy is a potentially salient feature for downstream Romanian NLP tasks such as coreference resolution, as it has been noted that this feature plays a role in the syn-

tax and semantics of Romanian (Dobrovie-Sorin, 1993). For this reason, automatic animacy classification of Romanian nouns is relevant not only for improving our understanding of the language, but also for future development of NLP tools that can accurately process Romanian text.

We present an approach to automatic classification of Romanian nouns by animacy, utilising a set of Romanian pretrained word embeddings (Păiș and Tufiş, 2018) and a seed set of word types with relevant animacy labels derived from Romanian WordNet (Dumitrescu et al., 2018). We hypothesize that information from these resources can provide an effective tool for Romanian noun classification with a relatively high degree of accuracy.

## 2. Background

The concepts of animacy and animacy classes in language and NLP have been defined in various ways. Zaenen et al. (2004) define animacy as a morphosyntactic feature that influences the grammatical prominence given to an entity in discourse. This prominence is determined based on three scales, namely animacy, definiteness and person, thus shaping how easily accessible entities are in language use. While these scales do not necessarily distinguish grammatical from ungrammatical use, they are rather useful in distinguishing felicitous from infelicitous language use. Therefore, even in languages with less complex morphology such as English, animacy plays a role in various processes. Examples include the selection of the genitive form or passive constructions, as well as the use of pronominal reference (Rosenbach, 2008). The semantic feature of animacy is also correlated with concepts such as agentivity and discourse salience. As Øvrelid (2005) notes, this is directly tied to markedness in discourse, as a subject noun is prototypically animate, which means that an inanimate noun in subject position

is an expression of a more marked structure. The importance of animacy information in discourse is further confirmed by event-related potential (ERP) studies that show this semantic feature is directly relevant from a processing perspective as well (Ji et al., 2020). Their study shows the importance of animacy in thematic and syntactic processing.

## 2.1. Animacy categories

Animacy is typically conceptualized as either a continuum, ranging from reference to a human entity, which is the ‘most animate’ end of the spectrum, to reference to an abstract inanimate entity (de Swart et al., 2008). Despite this continuous nature, it is typically divided into categories or a hierarchy of categories both in the linguistic literature and in the NLP literature. Natural language structures everything into categories, and the same applies to animacy. For example, languages divide the continuous space of perceivable colours into colour categories, each denoted by a basic colour term, and the inventory of basic colour terms differs between languages. The same applies to scales such as number (consider languages with a singular-dual-plural distinction), gender, individuation and also animacy. The extent to which the animacy of a nominal referent plays a role in its grammatical use tends to be divided into a very limited number of animacy categories, typically two. As noted by Bloem and Bouma (2013), ‘a linguistic phenomenon generally applies only to elements above a certain cut-off point in the hierarchy’ (p. 84). Dahl and Fraurud (1996) and Bayanati and Toivonen (2019) provide overviews with examples from diverse languages.

Consequently, in the field of NLP, animacy classification is also performed with various categorical classification schemes. Defining class labels is a crucial step in performing classification tasks based on animacy; at the same time, clear categorisation can be challenging due to the continuous nature of animacy. A binary classification choice between human and non-human classes seems to be the most common classification scheme in previous work. In this scheme, the non-human category does include various kinds of entities that we would consider animate to some extent but that are lower on the animacy scale than humans, such as animals, mythical creatures, autonomous agents, vehicles or natural forces. The human-nonhuman distinction seems to be a common animacy distinction in natural languages, though we are not aware of any typological work that has surveyed the frequencies of different types of animacy distinctions across languages. For these reasons, and as we are not aware of any morphosyntactic phenomena that rely on a finer-grained distinction in

Romanian, we also adopt a binary animacy distinction (human/nonhuman) in this work.

## 2.2. Related work

Øvrelid (2005) explores the effectiveness of decision-tree classifiers for automatic categorisation of a predefined set of Norwegian nouns, by using a number of linguistically-motivated morphosyntactic features, which were extracted from a pre-annotated corpus of Norwegian. The classifier was able to efficiently classify unseen nouns, demonstrating an accuracy of 90%, thus suggesting that such an approach might provide an effective and generalisable way of automating animacy classification for nouns. However, it is important to note that such an approach assumes the existence of large-scale, pre-annotated corpora, a resource that is less likely to be available for under-resourced languages.

A similar, more recent approach to noun classification is described in Klenner and Göhring (2021), where a gold standard for the animacy classification of German nouns is introduced and machine learning models are then applied to this data for classification purposes. The authors take into consideration potential metonymic use of words, thus accounting for cases in which the referent of a noun is animate, despite the noun itself being seemingly ambiguous in the absence of context. However, this is achieved through the use of manually annotated data. In a similar task, Ji et al. (2020) propose a method to automatically identify gender and animacy knowledge relevant for person mention detection. Their approach involves extracting noun-gender and noun-animacy pair counts from large n-gram data, subsequently using the informative pairs to identify person mentions from raw text.

Bowman and Chopra (2012) propose an approach based on a finer-grained animacy hierarchy of ten classes, by training a discriminative classifier on a corpus of spoken English that is pre-annotated with animacy labels. The classifier takes into account features that capture the internal structure of each noun phrase, as well as its syntactic relations to other key words in the sentence. The approach shows relatively high precision and recall on most classes and reaches an overall accuracy of 93.5% when the classes are merged for a binary (animate / inanimate) classification. One of the main reported errors made by the model involves classifying ambiguous cases where head words (often pronouns) can refer to entities belonging to various animacy classes, and no syntactic or morphological cue serves as indication to the correct one. This points to one of the main challenges of classifying into a fine-grained animacy hierarchy - when one is trying to predict animacy classes that do not correspond to the way animacy is gram-

matically expressed, there is greater potential for misclassification. In languages such as English, where different pronouns are used to refer to human (he/she) or non-human (it) entities, it is difficult to classify into more fine-grained classes unless the classifier can follow a co-reference chain to use features of the entity that is being referred to (Jahan et al., 2018). Conversely, animacy is also a useful input feature for the coreference resolution task (van Cranenburgh, 2019).

A possible account for classifying beyond grammatical animacy is presented by Evans and Orăşan (2000), who provide an example of using WordNet hierarchies as a lexical resource to improve classification accuracy. The described system incorporates this approach to account for the fact that there is generally little information about the form of common noun phrases that can point to the animate or non-animate status of the noun. As such, some level of world knowledge is required in order to make this classification and WordNet hierarchies can make up for this knowledge gap. The authors also incorporate this approach to animacy classification into an anaphora resolution system. Bloem and Bouma (2013) introduce the first animacy classification tool for Dutch using a combination of a seed set of noun types that were assigned an animacy label based on the Cornetto lexical-semantic database (a resource similar to WordNet, Vossen et al., 2007), and type-based classification using distributional features. They classify nouns into three animacy classes (human, nonhuman animate and inanimate), achieving a 93% accuracy – with some difficulty distinguishing the nonhuman category. A simplification to a two-class scheme (human / non-human), argued to be the most relevant distinction for Dutch, resulted in a 97% accuracy using a k-nearest neighbour algorithm.

We are not aware of any animacy classification work on languages that are more under-resourced than Romanian.

These studies reveal that resource availability strongly affects successful animacy classification. The approaches of Øvrelid (2005), Øvrelid (2009), Evans and Orăşan (2000) or Klenner and Göhring (2021) achieve high accuracy while involving the use of large-scale annotated corpora or lexical resources such as WordNet. This may pose replicability issues when attempting to generalise across languages, due to the fact that under-resourced languages might not benefit from the same lexical resources assumed by such approaches.

### 3. Methods

Our proposed classifier distinguishes between two classes of Romanian nouns, human and non-

human, by working with lemmas from a word list as opposed to tokens from a large annotated corpus. This entails that potentially ambiguous cases, specifically cases in which the same noun can refer to two different entities with different animacy statuses, will not be accounted for. We use a seed set of nouns labeled with animacy information derived from Romanian WordNet and make use of the associations encoded in a pretrained word embedding model (Păiș and Tufiș, 2018) to train a classifier that can generalize beyond the labeled seed nouns.

We first derived two sets of Romanian nouns from WordNet, attempting to encompass as many tokens that can be labelled as either human or non-human. To this end, we first identified two high-order hypernyms in the WordNet hierarchy that can act as either human or non-human targets for each subsequent set of words (namely *ființă umană* ‘human being’ and *artefact* ‘artefact’). These are the highest-order hypernyms in the WordNet hierarchy that should not contain instances of the opposite class. *Artefact* does have the hypernym of *entitate* ‘entity’ which covers even more non-human entities, but this synset also includes living entities as hyponyms. Including this would be prone to blurring the boundaries we defined for our classes, creating overlap, and potentially negatively affect classification accuracy.

After establishing these target high-order synsets, we used Open Multilingual WordNet (OMW)<sup>1</sup> to generate lists of hyponyms for each of the two synsets. This yields two different sets of words, one containing nouns with the semantic feature [+Human] and the other containing the semantic feature [+Non-Human]. This generation process simply consisted of taking all hyponyms of the high-order target synsets in the hierarchy from only the Romanian part of OMW.

Secondly, we extracted corresponding vectors for each word in this seed set using a set of pre-trained word embeddings for Romanian (Păiș and Tufiș, 2018). We use static embeddings as we are performing contextless type-based classification, and thus there is no context for contextual word embeddings to take advantage of. It is possible to distill static embeddings from contextual embedding models (Bommasani et al., 2020), but it is not clear whether this outperforms static approaches (Ehrmanntraut et al., 2021), especially for under-resourced languages.

The vectors are trained on text data from a reference corpus of Romanian (CoRoLa, Mititelu et al., 2014) using the Word2Vec skip-gram algorithm (Mikolov et al., 2013). The CoRoLa covers a wide array of functional language styles and domains,

<sup>1</sup>Open Multilingual WordNet. Version 1.6. <https://compling.hss.ntu.edu.sg/omw/>. Accessed: January 2023

including legal, scientific, journalistic, literary, and administrative, making it one of the largest fully IPR-cleared language corpora in the world. Out of the available pre-trained models, which differ in terms of hyperparameters, we used the model with a dimensionality of 300 and minimum type frequency of 20.

We subsequently incorporated the animacy information extracted from OMW into the word embeddings to create labeled feature vectors representing each noun in the dataset. We applied three different classifiers to the task: Random Forest (RF)<sup>2</sup>, Multi-layer Perceptron (MLP)<sup>3</sup> and k-nearest neighbours (KNN)<sup>4</sup>.

The RF classifier was chosen for its robustness to class imbalance, as typically the non-human or inanimate class is by far the most common, at least in terms of word type frequency and the number of WordNet sysets. The MLP classifier typically shows good performance for complex decision boundaries. The KNN classifier was also used by [Bloem and Bouma \(2013\)](#) for the Dutch animacy classification task under similar resource conditions.

We perform two evaluations with different aims. Firstly, we split the seed set of labeled nouns into a training and test set with an 80/20 split, and evaluate the accuracy on unseen noun types. However, this type-based evaluation is not a very realistic evaluation of animacy classification in a real-world context. In a corpus, the same noun type may belong to different animacy classes depending on context. Even though our classifier is type-based and thus does not take context into account when applied to a text corpus, we still perform a token-based evaluation of the classifier on natural language text in order to what extent this limitation of being type-based affects real-world animacy classification accuracy. Previous work often omits a token-based evaluation, as a type-based classifier will have a low performance ceiling on this task – we decided to include it in the interest of transparency and to reflect more realistic use cases.

## 4. Results

To our knowledge, we present the first Romanian animacy classifier, thus we do not have any previous baseline to compare our results to. We perform a type-based evaluation based on the WordNet-derived animacy labels, as well as a token-based

<sup>2</sup>Our RF hyperparameters were 100 estimators, Gini impurity as the splitting criterion, and no maximum tree depth.

<sup>3</sup>Our MLP hyperparameters were a hidden layer size of 100,  $\alpha = 0.0001$  and a learning rate of 0.001.

<sup>4</sup>Our KNN hyperparameters were  $k = 5$  and uniform weighting

Model	Precision	Recall	Accuracy
RF	<b>90.3%</b>	83.3%	88.5%
MLP	88.1%	87.5%	<b>89.1%</b>
KNN	70.0%	82.7%	76.4%

Table 1: Evaluation metrics obtained for each of the three classifiers.

evaluation based on a manually animacy-annotated gold standard of tokens in their textual context.

### 4.1. Type-based evaluation

A number of lemmas extracted from OMW in the first step, based on hypernymy/hyponymy relations, were not found in the vocabulary of the Romanian pretrained word embedding model, therefore the classification includes only those tokens extracted from OMW for which the utilised word embedding resource provided a vector. Following this filtering step, our final set of Romanian nouns includes a total of 5735 lemmas, out of which 3091 have non-human referents and 2644 have human referents. This sets the baseline accuracy of the classifier at 54.04%, i.e. the accuracy score that can be obtained by classifying every noun as non-human, the majority class.

Table 1 shows the results obtained by fitting the data to each of the three investigated classifiers: RandomForest, Multi-layer Perceptron and K-nearest neighbours. For the RF classifier, 90.3% of the predicted human-referent nouns were correctly classified and 88.5% of the predictions were correct. Compared to the baseline, these results indicate that the classifier performed relatively well in identifying nouns with a human referent. The MLP classifier achieves slightly better accuracy overall with 89.1%, due to better recall. The KNN algorithm shows significantly lower scores.

The choice between RF and MLP is a tradeoff of precision versus accuracy. In the broader context of using animacy classifiers in more complex text processing tasks, it is likely to be more important to correctly identify nouns referring to human entities, which is the positive class, thus favouring the RF classifier’s higher precision.

### 4.2. Token-based evaluation

In order to further evaluate the model on unseen data and gain insight into the limitations of a type-based classification approach on naturalistic data (as opposed to lexical-semantic databases, dictionaries or other structured inventories), we manually annotated Romanian text for a token-based evaluation. We randomly selected and extracted fifteen Wikipedia articles in Romanian, excluding incomplete or textually poor pages. The chosen

articles covered a wide range of topics, encompassing biographies of writers and athletes, as well as medical, political, and geographical topics. Following text preprocessing, we part-of-speech tagged the articles using a Romanian POS-tagging model from Python's Spacy library ('ro\_core\_news\_sm', [Honnibal and Montani, 2017](#)). A total of 954 noun tokens were extracted, with 902 remaining after eliminating occurrences that were either incorrectly tokenized (e.g. cut-off words) or incorrectly tagged as nouns ( $n = 52$ ).

These nouns were then manually annotated by a native speaker of Romanian, who labelled them as referring to a human or non-human entity. Importantly, all annotations were contextually informed, reflecting and adhering to the polysemy and context-dependency of certain nouns. For example, *martor* was identified throughout the data set as referring to both 'eyewitness' and 'reference sample'; similarly, *regim* was found to represent both 'diet' and 'government'. This highlights the potential limitations of a type-based classification approach, which was nevertheless necessitated by the under-resourced setting of Romanian. The newly annotated nouns consisted of 87.2% non-human referent nouns, 7.3% human referent nouns, and 5.5% tokens marked as 'other class' and eliminated. This dataset thus displays a pronounced imbalance, which is representative of real-world data where the human-referent category is underrepresented in comparison to the non-human referent category.

We then applied the RF classifier to this data, which has previously demonstrated the highest suitability for this task, and obtained an accuracy score of 59.9%. The confusion matrix generated by comparing the predicted labels to the ground truth of the human annotation revealed that the model correctly identified 514 true negatives and 27 true positives out of all 909 data points. Notably, the classifier also exhibited 43 false negatives, thus erroneously classifying a proportion of the evaluation data as denoting human entities. This might be due to the lack of a realistic class imbalance in the training set, causing the classifier to overpredict the less common 'human' class.

## 5. Discussion

We introduce the first Romanian animacy classifier, specifically a type-based binary classification of Romanian nouns into the classes human/non-human, using word embeddings and animacy information derived from WordNet. This type-based approach performs well on a type-based evaluation task, with both a Random Forest classifier and Multi-layer Perceptron classifier yielding high accuracy scores, but underperforms on a more naturalistic token-based evaluation task where animacy labels for the same

semantically ambiguous word type may differ by context. The type-based approach is necessitated by a lack of resources that prevent the creation of a token-based classifier, yet it is inadequate for solving the classification problem at a token level. This result underscores the inherent challenges in animacy classification for under-resourced languages in particular.

The classifier draws information from two relatively strong resources available for Romanian. The pre-trained vectors used in this approach are based on one of the largest and most comprehensive corpora of Romanian, making them the most feasible static word embeddings resource that we are aware of. Higher precision scores might be achieved by performing manual (or semi-manual) correction of the two animacy classes derived from WordNet, thus ensuring a lower error rate in labelling the training data. Alternatively, it might prove useful to explore a more fine-grained animacy hierarchy in classification in future attempts, such as the human/non-human animate/inanimate scale adopted by [Bloem and Bouma \(2013\)](#), although it needs to be taken into consideration to what extent a finer distinction in animacy has an impact on Romanian grammar or felicity.

Given that the main obstacle to good animacy classification for Romanian in a naturalistic setting is the type-based nature of our current approach, future work in Romanian animacy classification should explore the use of contextual embedding models such as Romanian BERT ([Dumitrescu et al., 2020](#)) for token-based classification where contextual representations can help to disambiguate animacy. Another line of future work would be to incorporate automatically classified animacy as a feature into relevant downstream NLP tasks for Romanian, such as coreference resolution.

## 6. Ethical considerations and limitations

As our token-based evaluation showed, the accuracy of our approach in a naturalistic setting is limited. Furthermore, in the token-based evaluation, we were not able to classify the nouns that are not present in the embedding model's vocabulary. This limitation can be addressed by using models that make use of subtoken embeddings if such a resource is available for the target under-resourced language, such as Romanian BERT ([Dumitrescu et al., 2020](#)) for the case of Romanian.

The reliance on the existence of a WordNet and pre-trained word embeddings for this language means our method cannot be extended to under-resourced languages that lack such resources.

As far as we are aware, the pre-trained embeddings that we used were not tested for the pres-

ence of harmful biases. It has been established that word embedding models often contain such biases, which could possibly lead to our classifier objectifying protected categories of human referents by misclassifying words that refer to them as non-human.

## 7. Bibliographical References

- Shiva Bayanati and Ida Toivonen. 2019. Humans, animals, things and animacy. *Open Linguistics*, 5(1):156–170.
- Jelke Bloem and Gosse Bouma. 2013. Automatic animacy classification for Dutch. *Computational Linguistics in the Netherlands*, 3:82–102.
- Rishi Bommasani, Kelly Davis, and Claire Cardie. 2020. Interpreting pretrained contextualized representations via reductions to static embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4758–4781.
- Samuel Bowman and Harshit Chopra. 2012. Automatic animacy classification. In *Proceedings of the NAACL HLT 2012 Student Research Workshop*, pages 7–10, Montreal, Canada. Association for Computational Linguistics.
- Osten Dahl and Kari Fraurud. 1996. Animacy in grammar and discourse. *Pragmatics and Beyond New Series*, pages 47–64.
- Peter de Swart, Monique Lamers, and Sander Lestrade. 2008. Animacy, argument structure, and argument encoding. *Lingua*, 118(2):131–140.
- Carmen Dobrovie-Sorin. 1993. *The Syntax of Romanian: Comparative Studies in Romance*. Walter de Gruyter.
- Daniel Stefan Dumitrescu, Andrei Marius Avram, Luciana Morogan, and Stefan-Adrian Toma. 2018. RoWordNet – A Python API for the Romanian Wordnet. In *10th International Conference on Electronics, Computers and Artificial Intelligence*, pages 1–6, Iasi, Romania.
- Stefan Dumitrescu, Andrei-Marius Avram, and Sampo Pyysalo. 2020. The birth of Romanian BERT. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4324–4328.
- Anton Ehrmanntraut, Thora Hagen, Leonard Konle, and Fotis Jannidis. 2021. Type-and token-based word embeddings in the digital humanities. In *CHR*, pages 16–38.
- Richard Evans and Constantin Orășan. 2000. Improving anaphore resolution by identifying animate entities in texts. In *Proceedings of the Discourse Anaphora and Reference Resolution Conference*.
- M Honnibal and I Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 688–697.
- Labiba Jahan, Geeticka Chauhan, and Mark Finlayson. 2018. A new approach to animacy detection. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1–12.
- H. Ji, S. Qi, S. Xu, J. Chen, D. Y. Dai, Y. Li, and W. Hu. 2020. The role of animacy in metaphor processing of Mandarin Chinese: An event-related potential (ERP) study. *Journal of Neurolinguistics*, 56.
- Manfred Klenner and Anne Göhring. 2021. The Detection of Actors for German. In *Text, Speech, and Dialogue*, pages 105–110, Cham. Springer International Publishing.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.
- Verginica Barbu Mititelu, Elena Irimia, and Dan Tufis. 2014. CoRoLa — the reference corpus of contemporary Romanian language. In *LREC*, pages 1235–1239.
- A. Pereltsvaig and E. Battistella. 2006. Gender and animacy in Russian: The interaction of morphology, semantics, and syntax. *Journal of Slavic Linguistics*, 14(1):49–90.
- Vasile Păiș and Dan Tufiș. 2018. More Romanian word embeddings from the RETEROM project. In *Proceedings of the International Conference on Linguistic Resources and Tools for Processing Romanian Language - CONSILR*, pages 91–100.
- Anette Rosenbach. 2008. Animacy and grammatical variation. Findings from English genitive variation. *Lingua*, 118(2):151–171.
- V. Sanchez-Gomez, E. Martinez-Camra, and L.A. Urena-Lopez. 2011. An empirical study on Spanish animacy classification for machine translation. In *Proceedings of the 13th International Conference on Computational Linguistics and Intelligent Text Processing*, pages 193–204.

Sapan Shah, Sreedhar Reddy, and Pushpak Bhat-tacharyya. 2020. *A retrofitting model for incorporating semantic relations into word embeddings*. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1292–1298, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Andreas van Cranenburgh. 2019. A Dutch coreference resolution system with an evaluation on literary fiction. *Computational Linguistics in the Netherlands Journal*, 9:27–54.

Piek Vossen, Katja Hofmann, Maarten de Rijke, Erik Tjong Kim Sang, and Koen Deschacht. 2007. The Cornetto database: Architecture and user-scenarios. In *Dutch-Belgian Information Retrieval Workshop*.

Annie Zaenen, Jean Carletta, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, Catherine O'Connor, and Tom Wasow. 2004. Animacy Encoding in English: Why and How. volume *Proceedings of the Workshop on Discourse Annotation*, pages 118–125, Barcelona, Spain. Association for Computational Linguistics.

Lilja Øvrelid. 2005. Animacy classification based on morphosyntactic corpus frequencies: some experiments with Norwegian nouns. In *Proceedings of the Workshop on Exploring Syntactically Annotated Corpora*, Birmingham.

Lilja Øvrelid. 2009. Empirical evaluations of animacy annotation. volume *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 630–638, Greece. Association for Computational Linguistics.

## A. Code and dataset

The following link provides access to the repository containing the code for our classification approaches, as well as our token-based evaluation dataset: <https://github.com/mariatepei/RO-animacy>