The Morpho-syntactic Annotation of Animacy for a Dependency Parser

Mohammed Attia, Vitaly Nikolaev, Ali Elkahky

Google Inc.

New York, USA

{attia,vitalyn,alielkahky}@google.com

Abstract

In this paper we present the annotation scheme and parser results of the animacy feature in Russian and Arabic, two morphologically-rich languages, in the spirit of the universal dependency framework (McDonald et al., 2013; de Marneffe et al., 2014). We explain the animacy hierarchies in both languages and make the case for the existence of five animacy types. We train a morphological analyzer on the annotated data and the results show a prediction f-measure for animacy of 95.39% for Russian and 92.71% for Arabic. We also use animacy along with other morphological tags as features to train a dependency parser, and the results show a slight improvement gained from animacy. We compare the impact of animacy on improving the dependency parser to other features found in nouns, namely, 'gender', 'number', and 'case'. To our knowledge this is the first contrastive study of the impact of morphological features on the accuracy of a transition parser. A portion of our data (1,000 sentences for Arabic and Russian each, along with other languages) annotated according to the scheme described in this paper is made publicly available (https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-1983) as part of the CoNLL 2017 Shared Task on Multilingual Parsing (Zeman et al., 2017).

Keywords: animacy, Arabic, Russian, Dependency Parsing

1. Introduction

The explicit encoding and identification of the animacy of entities in data has been reported to improve NLP tasks such as syntactic function disambiguation (Øvrelid, 2004; Lamers, 2007), anaphora resolution (Orasan and Evans, 2007a; Liang and Wu, 2004; Singh et al., 2014), syntactic parsing (Marton et al., 2011; Ambati et al., 2010; Nivre et al., 2008; Bharati et al., 2008), and disambiguation accuracy in boosting fluency in text generation (Bloem and Bouma, 2013).

In human cognition research, animacy is considered as one of the first characterizations made by human beings in their infancy and the last distinction lost in adults with Alzheimer's disease (Szewczyk and Schriefers, 2010). This is probably why computational representation of languages (particularly where animacy plays a morpho-syntactic function, e.g. Russian, Arabic, and Hindi) needs to provide adequate description and annotation of this feature. Contrary to previously prevalent perception that animacy is merely a semantic feature, recent years have seen increasing research showing animacy as permeating all levels of linguistic representations: morphology, syntax, semantics, and discourse (Folli and Harley, 2008; Ramchand, 2008; Ritter, 2014; Szewczyk and Schriefers, 2010). Moreover, animacy annotation for some languages is important for both downstream tasks as well as generation.

Animacy is a linguistic property that impacts and relates to a number of other linguistic phenomena, such as case marking, agreement, topicality, argument realization, and structural preferences. It has received formal linguistic investigation in the works of various researchers, most notably (Silverstein, 1976; Dik, 1997; Aissen, 2003; Martin et al., 2005; Dahl, 2008; Bloem and Bouma, 2013; Eckhoff, 2015; Karsdorp et al., 2015).

Various projects have targeted animacy annotation, most remarkably was the benchmark initiative of (Zaenen et al., 2004) where animacy annotation was included in two projects: the Paraphrase project and Possessive Alternation. The GNOME project (Poesio, 2004) aimed to create a corpus to study aspects of discourse and included an animacy taxonomy. The Dutch Cornetto lexical-semantic database (Martin et al., 2005) includes animacy annotation using hierarchical division of categories and subcategories. Orasan and Evans (2007a) described animacy annotation meant for anaphora resolution in English. Thuilier and Danlos (2012) annotated a French corpus for animacy and verb semantic classes. Jena et al. (2013) reported on work to enrich an already available treebank for Hindi with animacy information. Alkuhlani and Habash (2011), Elghamry et al. (2008), and Diab et al. (2014) reported on annotating Arabic data for gender, number and rationality (a hyponym of animacy).

1.1. Animacy Hierarchies

Different hierarchies and annotation schemes have been developed for animacy with different levels of granularity, but the common denominator in these hierarchies is the three level taxonomy first proposed by (Silverstein, 1976):

human > animate > inanimate

More fine-grained details are added to this core hierarchy by classifying 'human' according to person (1st, 2nd, and 3rd person) (Dik, 1997), breaking down 'inanimate' into 'concrete' and 'abstract' (Martin et al., 2005), classifying animates into 'organizations', 'animals', 'intelligent machines' and 'vehicles' (Zaenen et al., 2004) and making a special class for 'inanimate forces' (Dik, 1997) or 'natural forces' (DeLancey, 1981). This classification is assumed to emanate from the speakers' view of objects in the universe where humans are considered as more interesting and valuable than animals, and animals more so than things (Ransom, 1977).

For languages that employ animacy in their morpho-

syntactic paradigm, the exact meaning of animacy is also a source of variation from one language to the other. This is clearly exemplified by the comparison between animacy in Arabic and Russian. Russian makes distinction based on the lower scale of Silverstein's animacy hierarchy (animate vs. inanimate), while Arabic, makes the distinction on the higher scale, rational vs. irrational, which is the equivalent of human vs. non-human. In effect, the binary opposition in the animacy space in the two languages is carved out differently. The paradigms of animacy in Russian and Arabic are illustrated in Figure 1 where animacy in Russian is represented by the solid lines and in Arabic by the dotted ones.
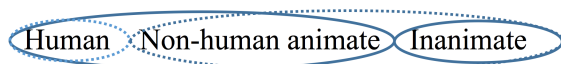

Figure 1: Comparing the notions of Animacy in Russian and Arabic

It is to be noted that animacy as a morpho-syntactic feature does not have to align with biological or ontological animacy. In biology, animacy is used to refer to living things, or anything that has a form of life, including animals and plants. Animals by contrast have the power of locomotion, which separates them from plants. Linguistic animacy is also not equivalent to the notion of 'living things', but is similar to the biological class of 'animal', yet modified by augmenting two additional features: sentience and volition (Jena et al., 2013), that are entities with intrinsic feeling and free will. This allows linguistic animacy to optionally exclude germs, viruses, and lower animates that are seen as collective entities, like cattle and insects. Moreover, what separates humans from non-humans is the feature of 'rationality'. This allows languages to optionally treat entities from outside homo sapiens, like devils and angels, as human.

We also note that animacy as a morpho-syntactic feature, does not have a universal definition, as there are exceptions and a certain element of arbitrariness in the assignment of this feature in different languages. In a survey of animacy in Penobscot, Quinn (2001) points out some unusual examples of nouns treated as animates, including nouns denoting fluid containers: kettle, pot and cup.

1.2. Animacy Types

There is a considerable amount of controversy regarding whether animacy is inside or outside of the "narrow syntax" (Folli and Harley, 2008; Ramchand, 2008; Ritter, 2014; Szewczyk and Schriefers, 2010), or in other terms whether animacy is a semantic/ontological category or a morpho-syntactic feature. We assume that the resolution to this question is to distinguish between the various types of animacy depending on how it is manifested across languages. (Wiltschko and Ritter, 2014) argued for the existence of two types of animacy: morphological animacy and high (semantic) animacy. We take this distinction further by hypothesizing three more types of animacy: anaphoric animacy, syntactic animacy and discourse animacy, adding up to a total of five types of animacy as explained below. These

types are not dichotomous, and the existence of animacy at one level, usually has relevance to animacy in the other levels.

Morphological animacy is a property of lexical entries and can additionally be featured in the inflection or derivation of words where lexemes at a certain animacy cutoff limit receive specific inflectional or derivational paradigms. In some languages animacy is a determinant of inflectional paradigms, like plural suffixes in Blackfoot, where animate nouns take the plural suffix -iksi while inanimates take the suffix -stsi (Wiltschko and Ritter, 2014), the plural suffix in the Sistani dialect of Persian (Shariphia et al., 2014) where -hā is used for non-human and ān for human nouns, and Arabic sound (regular) masculine plural suffix -uwna which is used exclusively with rational (human) entities. The derivational paradigm is exemplified by the Arabic collective nouns for lower animates (جراد (jarAd) – 'locusts', ذباب (*ubAb) – 'flies', نمل (namol) – 'ant') and lower inanimates (برتقال (burotuqAl) – 'oranges', تفاح (tufAH) – 'apple', and عنب (Einab) – 'grapes') where, contrary to other derivational tenancies, the singular is derived from the plural form.

Syntactic animacy can show in the grammar of a languages as strict rules and constraints or merely preferences. As a set of constraints, animacy is featured in the differential case assignment or gender-number agreement which is delineated by animacy, as detailed further in Section 2.. Languages which show inherent or explicit morphological animacy are expected to have syntactic animacy employed in one form or the other. Swart et al. (2008) points out that nouns that lie in the borderline between animate and inanimate may behave syntactically in both ways, reflecting the different aspects of their semantics (Jena et al., 2013). This is clearly exemplified by collective human nouns denoting organizations (such as tribe, team, police) which can be treated syntactically with alternate animacy agreement depending on whether the usage is literal or metonymical. Ambiguous words, like 'monster', and figurative language, like 'star' can also be considered on the borderline of the animacy divides, allowing them to optionally attach to different animacy classes.

As a set of preferences, animacy is assumed to affect the choice of the genitive type in English, where animate nouns tend to attract the *s*-genitive and inanimate nouns the *of*-genitive (Hinrichs and Szmrecsanyi, 2007; Altenberg, 1982).

Anaphoric (or pronominal) animacy is manifested in the animacy-based opposition in the pronoun system (Orasan and Evans, 2007b). English is a good example of this type of animacy. In English the pronouns 'he', 'she', 'it', 'who', and 'which' are demarcated along animacy (and gender) borderlines, and hence animacy is a property of referents, not of the morphology of the lexical items themselves. This feature lies on the border between felicity and grammaticality, depending on whether you see a phrase like 'The book and his reader' as ungrammatical or just unacceptable.

Semantic animacy is responsible for placing selectional restrictions on argument positions (Dahl and Fraurud,

1996). While the first three types of animacy are language-specific, semantic animacy is universal across all languages. It is the encyclopedic knowledge which determines, for instance, that the subject of the verb 'read' must be a human at a certain age. This feature determines sentence felicity, rather than grammaticality.

Discourse animacy is related to the effect of animacy (as an accessibility scale) on topicality and the salience of entities in texts. It is suggested that animacy plays an important role in determining entity prominence and how likely they are to be pronominalized (Poesio, 2004; McGill, 2009; Prat-Sala and Branigan, 2002).

The five types of animacy illustrate just how profoundly animacy permeates human language. This lends support to Dahl (2000)'s claim that animacy is "so pervasive in the grammars of human languages that it tends to be taken for granted and become invisible," and to Øvrelid (2004) 's statement that there is a "strong correlation between the animacy dimension and other linguistic dimensions." Nonetheless, the encoding of animacy in morphology and syntax is directly expressed in much fewer languages than the other features.

Our focus in this paper is on animacy as a purely morpho-syntactic phenomenon (i.e. the first two types above) where animacy applies to a certain, predefined cutoff limit. For example in Arabic and Russian, animacy has an immediate and tangible effect on the morphology and syntax of the two languages.

2. Animacy in Russian and Arabic Grammar

Despite the apparent differences in linguistic perception and dividing lines of animacy in Russian and Arabic (animate vs. inanimate and rational vs. irrational respectively) and despite the differences in the way animacy is featured in the syntax of both languages (differential object marking and differential gender-number agreement), we find that the two languages converge along many routes in the morphological and syntactic representation of animacy.

In morphology, the two languages meet along several demarcating lines when deciding the cutoff limits in animacy hierarchy. For example, groups of people are treated as inanimate in both languages(ru: армия (armiya) – ar: جيش (jayo$) – *'army'*; ru: народ (narod) – ar: شعب ($aEob) – *'nation'*). Groups of animals (ru: стадо (stado) – ar: قطيع (qaTiyE) – *'herd'*) and insects (ru: муравьи (muravyi) – ar: نمل (namol) – *'ants'*) are inanimate in Russian, and they receive inanimate morphology in Arabic by taking collective noun marking).

In syntax, both languages impose structural constraints based on animacy distinctions. In Arabic when a noun is both plural and irrational, it receives agreement equivalent to feminine gender and singular number, as shown by the following example.

الأولاد يلعبون (Al->awoladu yaloEabuwna.) 'The boys play.**pl.masc**'

القطط تلعب (Al-qiTaTu taloEabu.) 'The cats play.**sg.fem**'

In Russian animate nouns have identical inflection marking in genitive and accusative forms, while inanimates have identical inflection marking in nominative and accusative forms, as shown by the following example.

Я увидел мальчика. (Ya uvidel mal'chika.) 'I saw the boy.**gen**[1].'

Я читаю журнал. (Ya chitayu jurnal.) 'I am reading a journal.**nom**.'

3. Data Description

Over the past 25 years, the Linguistic Data Consortium (LDC) has been the main provider of annotated data and annotation standards. However, there are two main concerns with the LDC data (Marcus et al., 1995; Maamouri et al., 2003; Lander, 2005) that poses somes questions about its reliability and relevance to contemporary language. Firstly, most of the collected data is over two decades old now, which makes it less up-to-date. Interestingly, the last two decades have particularly witnessed many socio-political changes, information revolution, and many technological innovations leading to the coinage of many new terms and concepts. Secondly the LDC data is mainly focused on news edited texts, significantly limiting the robustness and scalability of parsing systems. The emergence of the social media and user-generated data have remarkably contributed to the surfacing of new or previously-ignored linguistic phenomena such as code-switching, informal texts and dialects, substandard spelling and grammatical constructions and the use of hashtags, repeated characters for emphasis, electronic addresses and emoticons.

As our annotation is eventually meant for aiding Information Retrieval, we focus on three document genres, namely, news, Wikipedia, and web documents. For the time being we do not include social media posts (e.g. from Facebook, Twitter or Google+) and we do not mine data from other genres, such as literature, religion, tourism, educational materials, technical manuals, spoken dialog, etc.

We annotated freshly collected data with roughly one-third from news articles (covering politics, sports, entertainment, business, health, sci-tech, arts), one-third from Wikipedia articles and one-third from web articles (including blogs, forums, reviews). We also try as much as possible to maintain the balance between sub-domains. This is to ensure that the data we collect is fresh, heterogeneous, application-oriented and representative.

Given the limited bandwidth of annotation time and cost, we also wanted to achieve the largest possible coverage with the smallest possible amount of sentences. Therefore we apply random sampling only for the initial 30% of the data. For the subsequent 70% we apply a 'word frequency-based sampling' approach which favors sentences with lexical items not seen in the initial set.

In this current work we annotate 10,466 sentences (189,029) for Russian and 9,717 sentences (399,774 to-

---

[1]We show the distinction between animate and inanimate paradigm by marking accusative case endings with the case label of the syncretic inflection. So for animate nouns accusative case is marked with **gen**, and for inanimate nouns the accusative case is marked with **nom**.

kens) for Arabic, with the average sentence length of 18.06 for Russian and 41.14 for Arabic. In the extracted data, we observed that only 0.39% of Russian sentence exceeded the 70 token limit while in Arabic (which exhibits the phenomenon of run-on sentences) we have 27.44% of the data going above that limit. Heuristically we decided to exclude all sentences exceeding 70 tokens.

4. Animacy Annotation Guidelines

Despite of the similarities between Russian and Arabic in the general perspective of animacy and its manifestation on linguistic representation, the two languages still have paradigmatic differences related to the cutoff limit within this feature (animate and inanimate vs. rational and irrational). This is why the annotation guidelines are developed independently to reflect the nature and idiosyncratic needs of each language.

4.1. Animacy in Russian Treebank Annotation

The Russian annotation scheme postulates that the category of animacy refers to the grammatical representation of nouns; it is a grammatical class and does not always align with the biological classification of living and nonliving entities. It affects the accusative case marking in singular forms of masculine and neuter nouns and the accusative case marking in plural forms of all nouns. It is also annotated on adjectival modifiers of these nouns.

Animate nouns of all three genders (feminine, masculine, neuter) have the same marking in the genitive and the accusative cases of the plural form. Likewise, all inanimate nouns regardless of gender have the same marking in the nominative and the accusative cases of the plural form. Unlike feminine nouns, masculine and neuter nouns show the same syncretic pattern across animacy also in the singular paradigm: animate nouns share inflections in the genitive and the accusative case and inanimate nouns share inflections in the nominative and the accusative cases. Animacy for feminine nouns and plural-only nouns are defined based on this pattern:

- Nominative_Singular: лампа "lamp"
- Nominative_Plural: лампы "lamps"
- Genitive_Plural: ламп "lamps"
- Accusative_Plural: лампы "lamps"= Nominative_Plural (inanimate)
- Nominative_Singular: мышь "mouse"
- Nominative_Plural: мыши "mice"
- Genitive_Plural: мышей "mice"
- Accusative_Plural: мышей "mice" = Genitive_Plural (animate)

The majority of neuter nouns is inanimate in Russian. There are several classes of exceptions, such as nouns referring to living creatures with the augmentative suffix -ище (-ische) like котище (kotische) – *'big cat'*, чудовище (chudovische) – *'monster'*, страшилище (strashilische) – *'ogre'*, etc. For non-masculine nouns, we identify the pattern of syncretism in plural paradigms to determine the animacy. If the accusative plural is identical to the genitive plural, we annotate the nouns as animate even in singular

forms where this morpho-syntactic distinction is not observed; likewise, if the accusative plural is identical to the nominative plural, we annotate inanimate.

4.2. Animacy in Arabic Treebank Annotation

The general principle of animacy annotation in Arabic is that all nouns (NNs) and proper nouns (NNPs) need to be tagged as either "rat" (rational), "irrat" (irrational) or "unps_a" (unspecified). Rationality mostly correlated with the state of humanness of the entity the noun denotes. The tag unsps_a is reserved for quantifiers (when used nominally), as they do not have intrinsic rationality and can be used variably to refer to rational or irrational objects, e.g. البعض Al-baEoD "some", الأكثر Al->akovar "most".

Two of the problem areas when annotating animacy in Arabic are pluralia tantum and metaphors and homonyms, and they are detailed below.

a) Pluralia Tantum

The pluralia tantum are special case plurals which break away from the productive methods of forming plurals from singular nouns. They refer to groups of people, animals or objects, and sometimes they are treated as units which can themselves be assigned plural inflection. They are of two main categories:

Group nouns. Plurals that can receive plural inflection, such as: جماعة jamAEap "group", فريق fariyq "team" (rationality: irrat).

Fixed plurals. These are plurals not derived from the singular forms, e.g. نساء nisA' "women", ناس nAs "people" (rationality: rat).

Mass nouns. These plurals that do not have singular forms, e.g. رمل ramol "sand", تراب turAb "dust", ضباب DabAb "cloud" (rationality: irrat).

Collective nouns. In this class, the lemma has a plural meaning, and the singular is derived from the plural by adding a morpheme (taa marboutah) in the end, such as: بقر baqar "cows", ذباب *ubAb "flies", تفاح tufAH "apples" (rationality: irrat).

b) Similes and Metaphors

Similes denote likeness between rational and irrational entities, while metaphors use an irrational entity to denote a rational one. In the former case, the human entity should be tagged as rational and the non-human entity as irrational. In the latter case, metaphors are treated as rational entities, as in the following example.

حضر الحفل العديد من النجوم HaDara Al-Hafola Al-Eadiydu mina Al-nujuwmi NN/rat
'Many stars NN/rat attended the party.'

5. Inter-Annotator Agreement Results

We computed inter-annotator agreement (IAA) for animacy (among other features) over a random set of 500 sentences. The purpose of this process is to determine if annotators require more training and/or if the guidelines need to be adjusted. The IAA corpus is representative of all document types and is annotated by different annotators in our teams, using the exact same workflow and guidelines. The IAA scores for Russian and Arabic are presented in Table 1.

Note that PoS is part-of-speech tagging accuracy, and LAS is Labeled Attachment Score (Buchholz and Marsi, 2006). The table shows that LAS scores are in the low 90s while PoS scores are in the high 90s. While aimacy IAA scores are above 90% for both Russian and Arabic, Russian scores are 5% higher, which indicate more error analysis is needed for Arabic.

Metric	Russian	Arabic
PoS	98.07	96.70
LAS	91.28	88.97
Per Token morphology accuracy	94.34	94.33
Animacy: F-measure	98.46	93.52
Number of annotators	5	5
Sentences annotated	500	454
Workflow	2-way	2-way

Table 1: IAA scores for Russian and Arabic

	Russian	Arabic
Total Sentences	3,985	3,745
total tokens	79,730	145,725
avg tokens per sentence	20.01	38.91
tokens with anim/rat	41,337	45,378
tokens with anim/rat %	51.85%	31.14%

Table 2: Percentage of tokens with animacy annotations

We also extract interesting statistics from the annotated data (full corpus). Table 2 shows the prevalence of the animacy feature in the annotated data. We also observe that the ratio of tokens annotated with animacy is different between Arabic (31.14%) and Russian (51.85%). This is due to the fact that the Arabic corpus has animacy annotations only for common and proper nouns, whereas the Russian corpus has animacy annotation on common and proper nouns, adjectives, participles, determiners, numerals, and several types of pronouns.

	NN-ru	NN-ar	NNP-ru	NNP-ar
anim/rat	13.82	9.93	44.27	45.55
inanim/irrat	86.18	90.07	55.73	53.59

Table 3: Animacy in NN and NNP

Table 3 demonstrates correlation between animate/rational and inanimate/irrational values in common (NN) and proper (NNP) nouns. Animacy in NNs and NNPs have a very similar distribution across Arabic and Russian where inanimate/irrational values are assigned to the majority of NNs, while we see a more balanced distribution of animacy in NNPs.

In regards to the correlation between gender and animacy, all genders across both languages show (Table 4) a general tendency towards a higher ratio of inanimate/irrational values associated with fem. This is contrasted by a more

	Russian %	Arabic %
fem anim/rat	7.57	5.81
fem inan/irrat	92.43	94.19
masc anim/rat	27.25	20.27
masc inan/irrat	72.75	79.73
neut anim	0.29	-
neut inan	99.71	-

Table 4: Gender to animacy correlation

balanced distribution in masculine nouns: Arabic 20.27% rational vs. 79.73% irrational; Russian 27.25% animate vs. 72.75% inanimate. Almost all Russian nouns with the neuter gender are inanimate.

In Table 5 we observe correlations between dependency function and animacy values in Russian and Arabic. For example, there is a general tendency for the animate/rational entities to occupy the subject position, and for the inananimate/irrational to occupy the position of an object of preposition. In Russian, 16.50% of animate nouns occupy the nsbuj position, compared to 7.09% of inanimate nouns. Similarly in Arabic, 18.77% of rational nouns are nsubj, compared to 9.16% of irrational nouns. Genitive modifiers (gmod) in both languages appear in the top three positions regardless of the animacy value.

6. Dependency Grammar

The dependency syntactic representation is a simple description of the grammatical relationships in a sentence. It represents a sentence as a set of binary asymmetrical dependency relations between each pair of tokens or constituents within the sentence. The dependencies merely express dominance relations between a governor (also known the head) and a dependent (also known as the child).

Dependency Grammar is usually viewed in contrast to Constituency Grammar (Chomsky, 1964) which represents sentences as phrase structure trees with a root node branching into non-terminal, terminal, and leaf nodes. It also uses empty nodes and traces for dropped elements. The advantage of Dependency Grammar is that it simplifies the representation by having one node for each token (or syntactic unit) and the arc is labelled with syntactic functions rather than phrase types. It does not employ the concept of empty nodes, making sure that the number of nodes corresponds to the number of tokens.

The origin of Dependency Grammars is usually traced back to the French grammarian Lucien Tesnière who is considered as the father of the theory (Osborne et al., 2015). Dependency Grammar subsequently evolved with contributions from a number of scholars most notably among them are Hays (Hays, 1964), Robinson (Robinson, 1970) and Mel'čuk (Mel'čuk, 1988). Computational implementation within the Dependency Grammars framework has been realized in the creation of dependency treebanks, such as the Prague Dependency Treebank (Hajič et al., 2001), the Stanford Dependencies (De Marneffe and Manning, 2008) and Universal Dependencies (Nivre et al., 2016; McDonald et al., 2013), and the development of dependency parsers,

ru inan %		ru anim %		ar irrat %		ar rat %	
23.23	amod	16.50	nsubj	36.19	pobj	22.52	nn
22.80	pobj	12.75	gmod	27.52	gmod	18.77	nsubj
14.25	gmod	11.83	amod	9.16	nsubj	16.32	gmod
7.09	nsubj	11.13	nn	8.21	conj	13.43	pobj
6.12	conj	9.19	conj	6.89	dobj	12.83	appos
5.10	dobj	8.85	appos	1.91	ROOT	7.03	conj
4.39	appos	8.01	pobj	1.41	appos	3.08	dobj
2.68	ROOT	4.40	ROOT	1.09	nn	1.13	attr
2.17	tmod	3.20	iobj	0.99	nsubjpass	0.89	ROOT

Table 5: Animacy and syntactic dependency

such as the Stanford parser (Chen and Manning, 2014), the inductive dependency parser (Nivre et al., 2004) and the MaltParser (Nivre et al., 2007).

Within Dependency Grammar, the dependency relations can be represented both in relational format and in graph format. In the relational format, the representation is a triple which shows a relation between a pair of words. The head of the dependency relation is given as the first argument and the dependent as the second. We represent this relation as follows:

```
relation(head, dependent)
```

For example, the sentence "he slept" can be represented as:

```
nsubj(slept, he)
```

which means "the nominal subject of 'slept' is 'he'", and the verb is the head of the pronoun.

Similarly, in the graph representation the dependency arc point from the head category to the dependent category, and the relation (or grammatical function) is represented as a label on the arc as shown in Figure 2.

- head-dependent relations (directed arcs)
- functional categories (arc labels)

Figure 2: Sample Dependency Graphs

7. Experimental Setup and Evaluation Results

For the prediction of animcy, in addition to the other morphological features, we train a linear-SVM classifier using features based on a window of the current word and word clusters of three words to the left and to the right, suffix for length 1, 2 and 3 for the current word, its left and right words, prefixes of lengths 1, 2 and 3 for the current word and the set of all known morphological attributes for the

current word paired with the observed POS tags in the training data.

In our experiments we use an arc-eager transition based dependency parser (Nivre, 2003) with a model trained using a linear SVM architecture similar to the one in Yamada and Matsumoto (2003). When experimenting with morphological features, we add the morphological attributes for both stack top and buffer top tokens.

7.1. Morphological Analysis

Apart from the morphological feature of animacy, our data is annotated for number, gender, case, aspect, mood, person, tense and voice, in addition to the Arabic-specific feature of definiteness and the Russian-specific features of number-antecedent and gender-antecedent.

The prediction results for animacy (as shown in Table 6) is 95.39% for Russian and 92.71% for Arabic which is slightly below the average (95.62% for Russian and 93.12% for Arabic).

	Russian %	Arabic %
animacy	95.39	92.71
number	97.60	96.37
gender	95.63	94.16
case	90.82	84.29
Average	95.62	93.12

Table 6: Evaluation results of morphological analysis

Tables 7 and 8 show the confusion matrices for 'animate' and 'inanimate' in Russian and 'rational' and 'irrational' for Arabic.

	Off-Diag	anim	inan
Off-Diagonal	197	268	589
anim	445	2,642	418
inan	396	226	14,906

Table 7: Confusion matrix for animacy in Russian

Table 9 shows the distribution of the sample space. It shows that in Russian and Arabic 'animate' and 'rational' are the minority classes (they occupy only 16% and 18%, less than one fifth of the sample space), and this correlates with the

2612

	Off-Diag	**irrat**	**rat**
Off-Diagonal	1030	1117	417
irrat	1,023	18,101	276
rat	681	425	3,581

Table 8: Confusion matrix for animacy in Arabic

high error rates for these classes (14.42% and 15.98% respectively). This shows that one possibility to improve the results is to treat the imbalance in the data, probably using active learning.

Feature	**Sample Total**	**Sample Ratio**	**Error Rate**
ru/anim	3,087	16.79%	14.42%
ru/inan	15,302	83.21%	2.59%
ar/rat	4,262	18.22%	15.98%
ar/irrat	19,124	81.78%	5.35%

Table 9: Animacy Sample Space and Error Rate

Setup	**Ru %**	**Loss**	**Ar %**	**Loss %**
All features	81.76	-	80.05	-
No animacy	81.01	0.75	79.95	0.10
No case	79.78	1.98	79.71	0.34
No gender	81.82	-0.06	79.91	0.14
No number	81.66	0.10	80.08	-0.03
No anim/case	79.79	1.97	79.78	0.27
No anim/gend	81.42	0.34	79.91	0.14
No anim/num	81.20	0.56	79.96	0.09
No anim/case num/gend	79.14	2.62	79.74	0.31

Table 10: Evaluation results of dependency parsing

7.2. Dependency Parsing

We conducted a number of experiment using different morphological tags as features for the transition parser to evaluate the effectiveness of animacy in improving the parser, and to compare animacy to other features found in nouns, independently and as a group or pairs.

Table 10 compares the parsing results using all morphological features to the situation where only one, a couple or a group of morphological features are removed. The results show that 'animacy' has a moderately positive impact on Russian and a slightly positive impact on Arabic. Compared to the other morphological features, 'animacy' is the second strongest morphological feature in Russian, while it ranks third among the four morphological features in Arabic.,

Table 10 also shows that morphological tags are helpful as features to the dependency parser, although the impact is much higher in Russian than in Arabic. Moreover, it is to be noted that the predicted tags (not the gold ones) are used

as features in the transition parser, and improving the morphological prediction accuracy would likely improve the dependency accuracy.

8. Conclusion

In this paper we have described the motivation for the annotation of animacy in a syntactic dependency treebank. We explained our annotation scheme for Russian and Arabic, two morphologically-rich, remotely related languages which have different views to the animacy scale. We have presented the results and statistics of the annotation of the animacy feature in our dependency treebanks for Russian and Arabic.

The annotated data is used to train a morphological analyzer and the results show a prediction f-measure of 95.39% and 92.71% for animacy in Russian and Arabic respectively. We also show that animacy along with other morphological tags can boost the performance of a dependency parser, and we assume that work on improving the morphological accuracy prediction can lead to further improvement on the parser.

9. Bibliographical References

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. In *Natural Language & Linguistic Theory, 21(3)*, pages 435–483.

Alkuhlani, S. and Habash, N. (2011). A corpus for modeling morpho-syntactic agreement in arabic: Gender, number and rationality. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: short papers*, pages 357–362, Portland, Oregon.

Altenberg, B. (1982). *The Genitive v. the Of-Construction. A Study of Syntactic Variation in 17th Century English.* CWK Gleerup.

Ambati, B. R., Husain, S., Nivre, J., and Sangal, R. (2010). On the role of morphosyntactic features in hindi dependency parsing. In *In Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages*, pages 94–102. Association for Computational Linguistics.

Bharati, A., Husain, S., Ambati, B., Jain, S., Sharma, D., and Sangal, R. (2008). Two semantic features make all the difference in parsing accuracy. In *Proc. of ICON 8*.

Bloem, J. and Bouma, G. (2013). Automatic animacy classification for dutch. In *Computational Linguistics in the Netherlands Journal 3*, pages 82–10.

Chen, D. and Manning, C. (2014). A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 740–750.

Chomsky, N. (1964). Aspects of the theory of syntax. Technical report, Massachusetts Institute of Technology.

Dahl, Ö. and Fraurud, K. (1996). Animacy in grammar and discourse. *Pragmatics and Beyond New Series*, pages 47–64.

Dahl, Ö. (2000). Animacy and the notion of semantic gender. In *In Gender in grammar and cognition: I: approaches to gender, ed. by Barbara Unterbeck, Matti Rissanen, Terttu Nevalainen and Mirja Saari*, pages 99–115. Berlin: de Gruyter.

Dahl, Ö. (2008). Animacy and egophoricity: Grammar, ontology and phylogeny. *Lingua*, 118:141–150.

De Marneffe, M.-C. and Manning, C. D. (2008). The stanford typed dependencies representation. In *COLING 2008: Proceedings of the Workshop on Cross-framework and Cross-domain Parser Evaluation*, pages 1–8. Association for Computational Linguistics.

de Marneffe, M.-C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J., and Manning, C. (2014). Universal stanford dependencies: A cross-linguistic typology. In *Proceedings of LREC-14*.

DeLancey, S. (1981). An interpretation of split ergativity and related patterns. pages 626–657. JSTOR.

Diab, M., Al-Badrashiny, M., Aminian, M., Attia, M., El-fardy, H., Habash, N., and Hawwari, A. (2014). Tharwa: A large scale dialectal arabic - standard arabic - english lexicon. In *The 9th edition of the Language Resources and Evaluation (LREC) Conference*, Reykjavik, Iceland.

Dik, S. C. (1997). The theory of functional grammar. In *Ed. Kees Hengeveld*. Berlin and New York: Mouton de Gruyter.

Eckhoff, H. M. (2015). Animacy and differential object marking in old church slavonic. *Russian Linguist*, 39.

Elghamry, K., Al-Sabbagh, R., and El-Zeiny, N. (2008). Cue-based bootstrapping of arabic semantic features. In *JADT 2008: 9es Journées internationales d'Analyse statistique des Données Textuelles*.

Folli, R. and Harley, H. (2008). Teleology and animacy in external arguments.

Hajič, J., Vidová-Hladká, B., and Pajas, P. (2001). The prague dependency treebank: Annotation structure and support. In *Proceedings of the IRCS Workshop on Linguistic Databases*, pages 105–114.

Hays, D. G. (1964). Dependency theory: A formalism and some observations. *Language*, 40(4):511–525.

Hinrichs, L. and Szmrecsanyi, B. (2007). Recent changes in the function and frequency of standard english genitive constructions: A multivariate analysis of tagged corpora. *English Language & Linguistics*, 11(3):437–474.

Jena, I., Bhat, R. A., Jain, S., and Sharma, D. M. (2013). Animacy annotation in the hindi treebank. In *Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse*, pages 159–167, Sofia, Bulgaria.

Karsdorp, F., Meulen, van der; Meder, M., Bosch, T., and van den, A. (2015). Animacy detection in stories. In *In Finlayson, M., Miller, B., Lieto, A. (ed.), Proceedings of the 6th Workshop on Computational Models of Narrative (CMN-2015)*, pages 82–97.

Lamers, M. J. (2007). Verb type, animacy and definiteness in grammatical function disambiguation. *Linguistics in the Netherlands*, 24:125–137.

Lander, T. (2005). Cslu: 22 languages corpus. In *LDC2005S26*, Philadelphia. Linguistic Data Consortium.

Liang, T. and Wu, D.-S. (2004). Automatic pronominal anaphora resolution in english texts. In *Computational Linguistics and Chinese Language*, pages 21–40.

Maamouri, M., Bies, A., Jin, H., and Buckwalter, T. (2003). Part 1 v 2.0. In *LDC2003T06*, Philadelphia. Linguistic Data Consortium.

Marcus, M., Santorini, B., and Marcinkiewicz, M. A. (1995). Treebank-2. In *LDC95T7*, Philadelphia. Linguistic Data Consortium.

Martin, W., Maks, I., Bopp, S., and Groot, M. (2005). Rbn-documentatie. In *Report, TST Centrale*.

Marton, Y., Habash, N., and Rambow, O. (2011). Improving arabic dependency parsing with form-based and functional morphological features. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 1586–1596, Portland, Oregon.

McDonald, R. T., Nivre, J., Quirmbach-Brundage, Y., Goldberg, Y., Das, D., Ganchev, K., Hall, K. B., Petrov, S., Zhang, H., Täckström, O., et al. (2013). Universal dependency annotation for multilingual parsing. In *The Annual Meeting of the Association for Computational Linguistics (ACL 2013)*, pages 92–97, Sofia, Bulgaria.

McGill, S. (2009). Gender and person agreement in cicipu discourse. In *PhD thesis*, School of Oriental and African Studies.

Mel'čuk, I. A. (1988). *Dependency syntax: theory and practice*. SUNY press.

Nivre, J., Hall, J., and Nilsson, J. (2004). Memory-based dependency parsing. In *Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004) at HLT-NAACL 2004*.

Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryigit, G., Kübler, S., Marinov, S., and Marsi, E. (2007). Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135.

Nivre, J., Boguslavsky, I. M., and Iomdin, L. L. (2008). Parsing the syntagrus treebank of russian. In *Proceedings of the 22nd International Conference on Computational Linguistics*. Association for Computational Linguistics.

Nivre, J., de Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C. D., McDonald, R. T., Petrov, S., Pyysalo, S., Silveira, N., et al. (2016). Universal dependencies v1: A multilingual treebank collection. In *International Conference on Language Resources and Evaluation (LREC 2016)*, pages 1659–1666, Portorož, Slovenia.

Nivre, J. (2003). An efficient algorithm for projective dependency parsing. In *Proceedings of the 8th International Workshop on Parsing Technologies (IWPT)*.

Orasan, C. and Evans, R. J. (2007a). Np animacy identification for anaphora resolution. In *Journal for Artificial Intelligence Research*, volume 29, pages 79–103.

Orasan, C. and Evans, R. J. (2007b). Np animacy identification for anaphora resolution. *Journal of Artificial Intelligence Research*, 29:79–103.

Osborne, T., Maxwell, D., and Center, U. C. (2015). A historical overview of the status of function words in dependency grammar. In *DepLing*, pages 241–250.

Øvrelid, L. (2004). Disambiguation of syntactic functions in norwegian: modeling variation in word order interpretations conditioned by animacy and definiteness. In

Proceedings of the 20th Scandinavian Conference of Linguistics, Helsinki: University of Helsinki.

Poesio, M. (2004). Discourse annotation and semantic annotation in the gnome corpus. In *The ACL 2004 Workshop on Discourse Annotation*, pages 72–79, Barcelona, Spain.

Prat-Sala, M. and Branigan, H. P. (2002). Discourse constraints on syntactic processing in language production: A cross-linguistic study in english and spanish. *Journal of Memory and Language*, 42:168–182.

Quinn, C. (2001). A preliminary survey of animacy categories in penobscot. In *Papers of the 32nd. Algonquian Conference*, pages 395–426.

Ramchand, G. (2008). In *Verb Meaning and the Lexicon: A First Phase Syntax*. Cambridge University Press.

Ransom, E. N. (1977). Definiteness, animacy, and np ordering. In *Proceedings of the 3rd Annual Meeting of the Berkeley Linguistics Society*, pages 418–429.

Ritter, E. (2014). Featuring animacy. In *Nordlyd 41.1:, special issue on Features edited by Martin Krämer, Sandra-Iulia. Ronai and Peter Svenonius. University of Tromsø*, pages 103–124. The Arctic University of Norway.

Robinson, J. J. (1970). Dependency structures and transformational rules. *Language*, pages 259–285.

Shariphia, S., Mirib, M., and Heidarypur, M. (2014). Investigation of relation of animacy and number in sistani dialect. *International Journal of Education and Research*, 2.

Silverstein, M. (1976). Hierarchy of features and ergativity. In *Dixon, R.M.W. (Ed.), Grammatical Categories in Australian Languages*, pages 112–171, Australian Institute of Aboriginal Studies, Canberra.

Singh, S., Lakhmani, P., Mathur, P., and Morwal, S. (2014). Analysis of anaphora resolution system for english language. *International Journal on Information Theory (IJIT)*, 3.

Swart, P. D., Lamers, M., and Lestrade, S. (2008). Animacy, argument structure, and argument encoding. *Lingua*, 118:131–140.

Szewczyk, J. M. and Schriefers, H. (2010). Is animacy special? erp correlates of semantic violations and animacy violations in sentence processing. *Brain Research*, 1368:208–221.

Thuilier, J. and Danlos, L. (2012). Semantic annotation of french corpora: animacy and verb semantic classes. In *LREC 2012*, pages 1533–1537. European Language Resources Association (ELRA).

Wiltschko, M. and Ritter, E. (2014). Animating the narrow syntax. In *Talk given at GLOW 37*.

Yamada, H. and Matsumoto, Y. (2003). Statistical dependency analysis with support vector machines. In *Proceedings of the 8th International Workshop on Parsing Technologies*, pages 195–206, Nancy, France.

Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitina, T., Catherine O'Connor, M., and Wasow, T. (2004). Animacy encoding in english: Why and how. In *Proceedings of the Workshop on Discourse Annotation*.

Zeman, D., Popel, M., Straka, M., Hajic, J., Nivre, J., Ginter, F., Luotolahti, J., Pyysalo, S., Petrov, S., Potthast, M., Tyers, F., Badmaeva, E., Gokirmak, M., Nedoluzhko, A., Cinková, S., jr, J. H., Hlaváčová, J., Kettnerová, V., Uresova, Z., Kanerva, J., Ojala, S., Missilä, A., Manning, C. D., Schuster, S., Reddy, S., Taji, D., Habash, N., Leung, H., de Marneffe, M.-C., Sanguinetti, M., Simi, M., Kanayama, H., Droganova, K., Alonso, H. M., Çağrı Çöltekin, Sulubacak, U., Uszkoreit, H., Macketanz, V., Burchardt, A., Harris, K., Marheinecke, K., Rehm, G., Kayadelen, T., Attia, M., Elkahky, A., Yu, Z., Pitler, E., Lertpradit, S., Mandl, M., Kirchner, J., Alcalde, H. F., Strnadová, J., Banerjee, E., Manurung, R., Stella, A., Shimada, A., Kwak, S., Mendonca, G., Lando, T., Nitisaroj, R., and Li, J. (2017). Conll 2017 shared task: Multilingual parsing from raw text to universal dependencies. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–19, Vancouver, Canada.