

Animacy Encoding in English: why and how

Annie Zaenen

PARC & Stanford University
3333 Coyote Hill Road
Palo Alto, CA 94304]
zaenen@parc.com

Jean Carletta

HCRC-University of Edinburgh
2, Buccleuch Place
Edinburgh EH8LW, UK
J.Carletta@Edinburgh.ac.uk

Gregory Garretson

Boston University
Program in Applied Linguistics
96 Cummington St.,
Boston, MA 02215
gregory@bu.edu

Joan Bresnan

CSLI-Stanford University
220, Panama Street
Stanford CA 94305
bresnan@stanford.edu

Andrew Koontz-Garboden

CSLI-Stanford University
220, Panama Street
Stanford CA 94305
andrewkg@csli.stanford.edu

Tatiana Nikitina

CSLI-Stanford University
220, Panama Street
Stanford CA 94305
tann@Stanford.edu

M. Catherine O'Connor

Boston University
Program in Applied Linguistics
96 Cummington St.,
Boston, MA 02215
mco@bu.edu

Tom Wasow

CSLI-Stanford University
220, Panama Street
Stanford CA 94305
wasow@csli.stanford.edu

Abstract

We report on two recent medium-scale initiatives annotating present day English corpora for animacy distinctions. We discuss the relevance of animacy for computational linguistics, specifically generation, the annotation categories used in the two studies and the interannotator reliability for one of the studies.

1 Introduction

It has long been known that animacy is an important category in syntactic and morphological natural language analysis. It is less evident that this dimension should play an important role in practical natural language processing. After reviewing some linguistic facts, we argue that it does play a role in natural language generation and translation, describe a schema for the annotation of animacy distinctions, evaluate the reliability of the scheme and discuss some results obtained. We conclude with some remarks on the importance of animacy and other accessibility dimensions for the architecture of generation schemes

2 The animacy dimension in natural language

The animacy hierarchy is one of the accessibility scales that are hypothesized to influence the grammatical prominence that is given to the realization of entities within a discourse. Three important scales (sometimes conflated into one, also called animacy hierarchy in Silverstein, 1976), are the definiteness, the person and the animacy hierarchy proper. We assume these are three different hierarchies that refer to different aspects

of entity representation within language: the definiteness dimension is linked to the status of the entity at a particular point in the discourse, the person hierarchy depends on the participants within the discourse, and the animacy status is an inherent characteristic of the entities referred to. Each of these aspects, however, orders entities on a scale that makes them more or less salient or 'accessible' when humans use their language.

The importance of accessibility scales is not widely recognized in computational treatments of natural language. This contrasts with the situation in linguistics where such scales have been recognized as playing an important role in the organization of sentence syntax and discourse.

Even in natural language studies, however, their importance has been underestimated because the role of these scales is not always to distinguish between grammatical and ungrammatical utterances but often that of distinguishing mainly between felicitous and infelicitous ones, especially in languages such as English.

Grammaticality and acceptability

As long as one's attention is limited to the distinction between grammatical and ungrammatical sentences, the importance of the animacy hierarchy in particular is mainly relevant for languages with a richer morphology than English. In such languages animacy distinctions can influence grammaticality of e.g. case-marking and voice selection. To give just one example, in Navaho, a *bi*-form is used rather than an *yi*-form

whenever the patient is animate and the agent is inanimate, whereas the *yi*-form is used when the agent is animate and the patient is inanimate as illustrated in (1) (from Comrie 1989 p. 193).

- (1) (a) At'éeéd nímasi *yi*-díílíd
girl potato burnt
The girl burnt the potato.
(b) At'éeéd nímasi *bi*-díílíd
girl potato burnt
The potato burnt the girl.

Other phenomena discussed in the literature are agreement limited to animate noun phrases (Comrie, 1989), overt case-marking of subject limited to inanimates or overt case-marking of objects limited to animates (Aissen, 2003, see also Bossong 1985 and 1991), object agreement in Bantu languages (see e.g. Woolford, 1999), choice between direct and inverse voice in Menominee (Bloomfield, 1962, see also, Trommer, n.d.).

Recent linguistic studies have highlighted the importance of the category in languages such as English. For instance the choice between the Saxon genitive and the *of*-genitive (Rosenbach, 2002, 2003, O'Connor et al. 2004, Leech et al. 1994), between the double NP and the prepositional dative (Cueni et al, work in progress) and between active and passive (Rosenbach et al. 2002, Bock et al. 1992, McDonald et al. 1993) and between pronominal and full noun reference (Dahl and Fraurud, 1996, based on Swedish data) have all been shown to be sensitive to the difference between animate and inanimate referents for the arguments with variable realization. In these cases the difference between animate and inanimate does not lead to a difference between a grammatical or an ungrammatical sentence as in the cases cited in the previous paragraph but to a difference in acceptability.

Interaction between animacy and other scales

As mentioned above, the term 'animacy hierarchy' is used in two ways, one to refer to an ordering imposed on definiteness, animacy and person and the other where it refers to animacy proper. The reason for this lies in the interaction between the different factors that determine accessibility.

It is conceptually desirable to distinguish between animacy and definiteness but in practice it is frequently the case that a linguistic phenomenon is conditioned by more than one of these conceptually independent factors (see e.g. Comrie, 1989 for discussion). The projects in the context of which the annotation tasks described here were

performed (see description below, section 5) also encode some of these interacting factors. The LINK-project encodes information status (see Nissim et al, 2004) and the Boston project encodes definiteness and expression type (i.e., NP form) as proxies for information status.

3 Animacy as a factor in generation and translation

As long as animacy was discussed as a relevant grammatical category in languages that had not been studied from a computational point of view, its importance for computational linguistics was perceived as rather limited. The fact that it permeates the choice between constructions in languages such as English changes this perception. The category is of obvious importance for high quality generation and translation.

For instance, if one is faced with the task of generating a sentence from a three place predicate, P (a,b,c), and one has the choice of rendering either b or c as the direct object, knowing that c is human and b is abstract would lead one to choose c *ceteris paribus*. However, everything else is rarely equal and the challenge for generation will be to assign the exact relative weights to factors such as animacy, person distinction and recency. Moreover, the importance of these factors needs to be combined with that of heterogeneous considerations such as the length of the resulting expression.

In the context of translation we need also to keep in mind the possibility that the details of the animacy ranking might be different from language to language and that the relative impact of animacy and other accessibility scale factors might be different from construction to construction.

4 The Animacy Hierarchy

Given the pervasive importance of animacy information in human languages one might expect it to be a well-understood linguistic category. Nothing could be farther from the truth. Linguistic descriptions appeal to a hierarchy that in its minimal form distinguishes between human, non-human animate and inanimate but can contain more distinctions, such as distinctions between higher and lower animals (see Yamamoto, 1999 for a particularly elaborated scheme).

What makes it difficult to develop clear categories of animacy is that the linguistically relevant distinctions between animate and non-animate and between human and non-human are not the same

as the biological distinctions. Part of this research is devoted to discovering the principles that underlie the distinctions; and the type of distinctions proposed depend on the assumptions that a researcher makes about the underlying motivation for them, e.g. as a reflection of the language user's empathy with living beings (e.g. Yamamoto, 1999). What is of particular interest for natural language processing is the observation that the distinctions are most likely not the same across languages (cf. Comrie, 1989) and can even change over time in a given language. They are similar to other scalar phenomena such as voicing onset times that play a role in different languages but where the categorization into voiced and unvoiced does not correspond to the same physical boundary in each language. But whereas voicing onset times can be physically measured, we do not have an objective measure of animacy. The categories involved correspond to the degree to which various entities are construed as human-like by a given group of speakers and at this point we have no language independent measure for this.

Moreover, languages make ample use of metaphor and metonymy. The intent of an animacy coding is to encode the animacy status of the referent of the linguistic expression. But sometimes in figurative language it is not clear what the referent is. Especially prevalent cases of metonymy are the use of names to refer both to organizations (e.g. IBM) and to characteristic members of them, and the use of place names (e.g. Russia) to refer both to organizational entities and geographical places or inhabitants of them. Terms belonging to these semantic classes are systematically ambiguous. Whereas it is true that animacy can be determined by looking at the entity an expression refers to, in practice it is not always clear what the referent of an expression is.

The notions that the animacy hierarchy appeals to, then, are not a priori well defined. And work is necessary on two levels: to better define which distinctions play a role in English and to determine where they play a role. Conceptually, it might be desirable to replace the idea of a hierarchy with discrete classes by a partial ordering of entities. This is, however, not the place to pursue this idea.

Fortunately, one doesn't need to wait until the first problem is solved completely to tackle the second. The results obtained in certain linguistic contexts are robust for the top and the bottom of the hierarchy. Uncertainty about the middle does not prevent us from establishing the importance of the dimension as such. Refining the definition of

animacy will, however, be important for more detailed studies of the interaction between the various accessibility hierarchies. This more precise notion will be needed for cross-linguistic studies, and, in the context of natural language processing, for high quality generation and translation.

5 Animacy Annotation

As we have discussed above, the animacy scale is an important factor in the choice of certain construction in English. But it is only a soft constraint and as such outside of the realm of things that native speakers have clear judgments about. The best ways to study such phenomena are psychological experiments and corpus analysis. The annotation exercise we engaged in is meant to facilitate the latter.

Given the situation described with respect to animacy categories, a natural way to proceed is to start with a commonsensical approach and see where it leads. In 2000-2002, two rather similar initiatives led to the need for animacy annotations: one, the paraphrase-project, a collaboration between Stanford and Edinburgh, concentrating on the factors that influence the choice between different sentence level syntactic paraphrases (Bresnan et al. 2002) and another concentrating on the possessive alternation (O'Connor, 2000). The two projects used a very similar animacy annotation scheme, developed in the context of the O'Connor project.

The scheme was used in two different ways. The Boston team coded 20,000 noun phrases in 'possessive' constructions from the Brown Corpus. The first round of coding was automated, with the animacy annotation based primarily on word lists and morphological information. The second round was performed manually by pairs of coders using a decision tree. The two coders were required to agree on each code; every case in which there was not complete agreement was discussed by the rest of the team, until a choice of code was made. This way of annotating does not lend itself to a study of reliability, except between the automated coder and the human coders as a group. For more information on this use of the coding system, see Garretson & O'Connor (2004).

In what follows we concentrate on the use of the coding scheme in the Stanford-Edinburgh paraphrase project.

The overall aim of the paraphrase project is to provide the community of linguists and computational linguists with a corpus that can be

used to calculate the impact of the various factors on different constructions. The annotation scheme assumes that the main distinction is three-way: human, other animates and inanimates, but the two latter categories are subdivided further as follows:

- Other animates: organizations, animals, intelligent machines and vehicles.
- Inanimates: concrete inanimate, non-concrete inanimate, place and time.

The category ‘organization’ is important because organizations are often presented as groups of humans engaging in actions that are typically associated with humans (they make pronouncements, decisions, etc.). The categories place and time are especially important for the possessive encoding as it has often been observed that some spatial and temporal expressions are realized as Saxon genitives (see e.g. Rosenbach (2002)).

For the cases in which no clear decision could be made, a category ‘variable animacy’ was invented, and the coders were also given the option to defer the decision by marking an item with ‘oanim’.

The overall coding scheme, with a summary of the instructions given to the coders, looks as follows¹

HUMAN

Refers to one or more humans; this includes imaginary entities that are presented as human, gods, elves, ghosts, etc.: things that look human and act like humans.

ORG

This tag was proposed for collectivities of humans when displaying some degree of group identity. The properties that are deemed relevant can be represented by the following implicational hierarchy:

- +/- chartered/official
- +/- temporally stable
- +/- collective voice/purpose
- +/- collective action
- +/- collective

The cut-off point between HUMAN and ORG was put at ‘having a collective voice/purpose’: so a group with collective voice and purpose is deemed to be an ORG, a group with collective action, such as a mob, is not an ORG.

ANIMAL

Non-human animates, including viruses and bacteria.

PLACE

The tag is used for nominals that ‘refer to a place as a place’. There are two different problems with the delimitation of place. On the one hand, any location can be a place, e.g. a table, a drawer, a pinhead, ... The coding scheme takes the view that only potential locations for humans are thought of as ‘places’. On the other hand some places can be thought of as ORGs. The tag was applied in a rather restricted way, for instance in a sentence such as ‘my house was built in 1960’, ‘my house’ is coded as CONC (see below), whereas in ‘I was at my house’, it would be a PLACE.

TIME

This tag is meant to be applied to expressions referring to periods of time. It was applied rather liberally.

CONCRETE

This tag is restricted to ‘prototypical’ concrete objects or substances. Excluded are things like air, voice, wind and other intangibles. Body parts are concrete.

NONCONC

This is the default category. It is used for events, and anything else that is not prototypically concrete but clearly inanimate.

MAC

A minor tag used for intelligent machines, such as computers or robots.

VEH

Another minor category used for vehicles as it has been observed that these are treated as living beings in some linguistic contexts (e.g. pronoun selection in languages such as English where normal gender distinctions only apply to animates).

OANIM

This tag is used when the coder is completely unsure and wants to come back to the example later.

VANIM

This tag can be used in conjunction with another one to indicate that the coder is not entirely sure of the code and thinks there are reasons to give another code too.

¹ For a more extensive description of the annotation scheme see Garretson et al. 2004.

Finally, NOT-UNDERSTOOD was supposed to be used when the text as a whole was not clear.

Three coders coded the parsed part of the Switchboard corpus (Godfrey et al. 1992) over the summer of 2003. The corpus consists of around 600 transcribed dialogues on various predetermined topics among speakers of American English. Before the annotation exercise began, the dialogues were converted into XML (Carletta et al. 2004). The entities that needed to be annotated (the NPs and possessives determiners) were automatically selected and filtered for the coders. The three coders were undergraduate students at Stanford University who were paid for the work. The schema presented above was discussed with them and presented in the form of a decision tree. Difficult cases were discussed but eventually each coder worked independently. 599 dialogues were annotated.

6 Coding reliability

The reliability of the annotation was evaluated using the kappa statistic (Carletta, 1996). Although there are no hard and fast rules about what makes an acceptable kappa coefficient—it depends on the use to which the data will be put—many researchers in the computational linguistics community have adopted the rule of thumb that discourse annotation should have a kappa of at least .8.

For the reliability study, we had three individuals work separately to code the same four dialogues with the animacy scheme. Markables (in this case NPs and possessives) had been extracted automatically from the data, leading the coders to mark around 10% of the overall set with a category that indicated that they were not proper markables and therefore not to be coded. Omitting these (non-) markables, for the data set overall, $K=.92$ ($k=3, N=1081$).

In general, coders did not agree about which cases were problematic enough to mark as VANIM, and omitting the markables that any coder marked as problematic using the VANIM code leads to a slight improvement ($K=.96, k=3, N=1135$).

It is important to note that these kappa coefficients are so high primarily because two categories which are easy to differentiate from each other, HUMAN and NONCONC, swamp the rest of the categories. The cross-coder reliability for them is satisfactory but the intermediate categories were not defined well enough to allow reliable coding.

Figure 1 shows the confusion matrix for the data including markables that any coder marked additionally as problematic using the VANIM notation. Considering the coders one pair at a time, the matrix records the frequency with which one coder of the pair chose the category named in the row header whilst the other chose the category named in the column header for the same markable.

Although we were aware of the less than formal definitions given for the categories, we had hoped that the coders would share the intuitive understanding of the developers of the categories. This is obviously not the case for all categories. What was also surprising was that allowing coders to mark cases as problematic using the VANIM code was not worthwhile, since the coders did not often take advantage of this option and taking the VANIM codes into account during analysis has little effect.

Analysis of the four annotated dialogues points to several sources for the intercoder disagreement.

- The categories TIME and PLACE were defined in a way that did not coincide with the coders' intuitive understanding of them. The tag TIME was supposed to refer to 'periods of time'. This led to some wavering interpretations for temporal expressions that do not designate a once-occurring period of time. For instance 'this time' and 'next time' were coded as TIME by two coders but as 'NONCONC' by the third one. Clearer training on what was meant could have helped here.
- As mentioned above, the choice between HUMAN, ORG and NONCONC depended on how the coders interpreted the referent of the expression. Although guidelines were given about the difference between HUMAN and ORG (see above), the cut-off point wasn't always clear². The distinction between ORGs as proposed in our schema and less organized human groups seems too fluctuating to be useful.

2 All coders agreed that Vulcans are HUMAN.

	concrete	human	non-conc	not-understood	oanim	oorg	place	time	veh
concrete	31	9	19	10	0	0	5	0	1
human		1489	27	11	0	33	0	4	0
nonconc			1256	3	3	33	23	53	1
notunderstood				0	0	0	0	0	0
oanim					2	2	0	1	0
org						91	0	0	0
place							66	0	2
time								70	0
veh									3

Figure 1

- The vagueness of pronominal reference: for instance a school as an organization can be marked as ORG by the coders but later in the dialogue there is discussion about the what is done with napping children in the school and one speaker says ‘if they (the children) fall asleep they kind of let them sleep’, one coder interpreted that the second ‘they’ as simply referring to the school organization and marked it as ORG, whereas another interpreted it as referring to a rather vague group of humans, presumably some teachers, and marked it as HUMAN. This vagueness of reference is quite prevalent in spoken language, especially with the pronoun ‘they’.
- Attention errors, e.g. vehicles were supposed to get a special code but, presumably because there were so few, this was sometimes forgotten. One coder coded ‘a couple of weeks’ as HUMAN. These kinds of mistakes are unavoidable and the very tools that make the encoding easier (e.g. the automatic advancing from one markable to another) might make them more frequent.

While the problems with ORG and HUMAN don’t come as a surprise, the difficulties with PLACE, TIME and CONCRETE are more surprising. The two minor classes, MAC and VEH and the ANIMAL class occurred so seldom that no significant results were obtained in this sample. They were equally rare in the corpus as a whole.

7 Conclusion

We are not aware of any other medium-scale attempts to annotate corpora of contemporary English for animacy information apart from the two mentioned here. There are smaller efforts concentrating on the genitive alternation (e.g. Leech et al., 1994, Anschutz, 1997, Stefanowitsch, 2000)³. The resources that have been created give robust results for the opposition ‘human’ versus ‘nonconcrete’ entities in the large sense (as the category was used as a catch-all). This should suffice for further inquiries about linguistic processes that are sensitive to a binary opposition in this dimension. Moreover the Stanford-Edinburgh effort is integrated in a corpus that has already been marked up for syntactic information, so correlations between syntactic constructions and animacy (and information status) should be easy to calculate. It is also the first effort that studies inter-annotator reliability.

Some studies based on the annotations are currently being conducted. The study by Cueni, Bresnan, Nikitina and Baayen (2004) supplements partial data from the work described here with further annotations. The work reported by O’Connor et al. (2004) derives from the Boston use of the encoding described here. Within the paraphrase project we are currently investigation

³ Some related work is done in the context of entity tracking sponsored by various US government programs (ACE, TIDES, etc.). The proposed annotation schemes have problems in distinguishing between persons and organizations or geo-political entities that are similar to ours, but the basic categories and the aims of these enterprises are different. We have not reviewed them here.

the possible effect of animacy on constructions such as Left-Dislocation and Topicalization.

Further work remains to be done, however, to determine the exact nature of the distinctions in the animacy dimension that are important for English and for other languages. The annotations we provide do not settle this issue. In that sense they are insufficient to guide generation and translation precisely. To investigate this further we will need to devise more careful annotation schemes and approach the problem via experiments where the hypothesized relative animacy of various entities can be carefully controlled. As mentioned above, it might be better not to think in terms of robust large categories but rather try to rank specific entities or small categories relative to each other and to gradually build up a more precise picture. This is most likely better done through controlled experiments than through corpus annotation.

The annotated corpus, however, will be helpful to determine where animacy plays a role and which other factors it interacts with. This knowledge will help devise more adequate generation and translation architectures.

The Boston University noun Phrase Corpus is publicly available at <http://np.corpus.bu.edu>. The paraphrase corpus will be made available to subscribers to the Switchboard Corpus.

8 Acknowledgements

This work was in part supported by a Scottish Enterprise Edinburgh-Stanford Link Grant (265000-3102-R36766) and by NSF grant BCS-0080377. Thanks to Toni Jeanine Harris, Rebecca Regos and Anna Cueni for the encoding work and to Richard Crouch and Neal Snider for comments and help. The usual disclaimers obtain.

References

- Aissen, Judith, 2002. Differential object marking: Iconicity vs. economy. *NLLT*, 21 435-483.
- Anschutz, A. 1997. "How to Choose a Possessive Noun Phrase Construction in Four Easy Steps." *Studies in Language* 21, 1, 1-35.
- Bock, J. K., Loebell, H. and Morey, R. (1992) From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review* 99: 150--171.
- Bossong, Georg. 1985. *Differentielle Objektmarkierung in den Neuiranischen Sprachen*, Gunter Narr Verlag, Tübingen.
- Bossong, Georg. 1991. 'Differential Object Marking in Romance and Beyond', in D. Wanner and D. Kibbee (eds.), *New Analyses in Romance Linguistics: Selected Papers from the XVIII Linguistic Symposium on Romance Languages*, John Benjamins, Amsterdam, pp. 143-170.
- Bresnan, Joan, Dingare, Shipra, and Manning, Christopher D. 2001. Soft Constraints Mirror Hard Constraints: Voice and Person in English and Lummi. [Proceedings of the LFG '01 Conference](#). CSLI Online. 20 pages.
- Bresnan, Joan, Carletta, Jean, Crouch, Richard, Nissim, Malvina, Steedman, Mark, Wasow, Tom and Zaenen, Annie. 2002. Paraphrase analysis for improved generation, LINK project, HCRC Edinburgh-CLSI Stanford.
- Carletta, Jean. 1996. Assessing agreement on classification task: the kappa statistic. *Computational linguistics*, 22 (2): 249-254
- Carletta, Jean, Shipra Dingare, Malvina Nissim, and Tatiana Nikitina. 2004. Using the NITE XML Toolkit on the Switchboard Corpus to study syntactic choice: a case study. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC2004), Lisbon, May 2004*.
- Comrie, Bernard, 1989, *Language Universals and Linguistic Typology*, The University of Chicago Press
- Cueni, Anna, Joan Bresnan, Tatiana Nikitina and Harald Baayen. 2004. Predicting the Dative Alteration, Stanford University, ms. in preparation.
- Dahl, Östen and Fraurud, Kari. 1996 Animacy in Grammar and Discourse. In Thorstein Fretheim & Jeanette K. Gundel (eds.), *Reference and Referent Accessibility*, 47-64. Amsterdam/Philadelphia: [John Benjamins](#).
- Garretson, Gregory, O'Connor, M. Catherine; Skarabela, Barbora; & Hogan, Marjorie. March 2004. *Coding practices used in the project Optimal Typology of Determiner Phrases*. <http://npcorpus.bu.edu/documentation/index.html>
- Garretson, Gregory, and O'Connor, M. Catherine. 2004. A combined automatic-and-manual method for studying discourse features in corpora. Paper to be presented at *the Fifth North American Symposium on Corpus Linguistics*, May 21-23, Montclair State University, NJ.
- Godfrey, J. Holliman, E. & McDaniel J. 1992 SWITCHBOARD: Telephone speech corpus for research and development. *Proceedings of ICASSP-92*, 517-520.
- Leech, G., B. Francis, et al. (1994). The Use of Computer Corpora in the Textual Demonstrability of Gradiance in Linguistic Categories. *Continuity In Linguistic Semantics*.

- C. Fuchs and B. Victorri. Amsterdam, John Benjamins Publishing Company: 57-76.
- McDonald, J.L., Bock, K. and M. Kelly. 1993. Word and world order: semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology* 25: 188-230.
- Nissum, Malvina, Shipra Dingare, Jean Carletta and Mark Steedman. 2004. An annotation scheme for information status in dialogue. Submitted to LREC 2004
- O'Connor, M. Catherine. 2000. Optimality typology of the DP: Markedness within the nominal. NSP-grant. BCS-0080377.
- O'Connor, M. Catherine., Anttila, Arto, Fong, Vivienne and Maling, Joan (2004). Differential possessor expression in English: Re-evaluating animacy and topicality effects. Paper presented at the Annual Meeting of the Linguistic Society of America, January 9-11, Boston, MA.
- Rosenbach, Anette (2002) *Genitive Variation in English. Conceptual Factors in Synchronic and Diachronic Studies*. Berlin/New York: Walter de Gruyter.)
- Rosenbach, Anette. 2003. Aspects of iconicity and economy in the choice between the *s*-genitive and the *of*-genitive in English. In: Günter Rohdenburg & Britta Mondorf (eds.). *Determinants of Grammatical Variation in English* Berlin/New York: de Gruyter.
- Rosenbach, A., J. Aissen, and J. Bresnan. 2002. Pilot study of the influence of animacy on subject choice in a reading task. (Heinrich Heine University, UCSC, and Stanford)
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Richard Dixon, editor, *Grammatical Categories in Australian Languages*. Australian Institute of Aboriginal Studies.
- Stefanowitsch, Anatol. 2000. Constructional semantics as a limit to grammatical alternation: The two genitives of English. in *CLEAR* (Cognitive Linguistics: Explorations, Applications, Research), 3.
- Trommer, Jochen. s.d. Direction Marking and Case in Menominee, <http://www.ling.uni-osnabrueck.de/trommer/nim.pdf>
- Woolford, Ellen. 1999. Animacy hierarchy effects on object agreement, in Paul Kotey ed. *New Dimensions in African Linguistics and Languages* (Trends in African Linguistics 3), 203-216.
- Yamamoto, Mutsumi, 1999, Animacy and Reference: a cognitive approach to corpus linguistics, John Benjamins.