Adjectives as saturators vs. modifiers: Statistical evidence

Gemma Boleda[1], Stefan Evert[2], Berit Gehrke[1], and Louise McNally[1]

[1] Universitat Pompeu Fabra, Roc Boronat 138,
08018 Barcelona, Spain
{gemma.boleda,berit.gehrke,louise.mcnally}@upf.edu,
http://www.upf.edu/
[2] Technische Universität Darmstadt, Hochschulstrasse 1,
64289 Darmstadt, Germany
evert@linglit.tu-darmstadt.de,
http://www.tu-darmstadt.de/

Abstract. This paper reports on a large-scale, statistical analysis of corpus data to support the null hypothesis that ethnic adjectives (EAs, e.g. *French*) are ordinary adjectives, rather than argument-saturating "nouns in disguise" (in, e.g., *French agreement*). In particular, EAs are argued to simply modify the noun they combine with; their special properties in inducing argument-like behavior arises from the interaction between the semantics of event nominals and that of the adjective.

Keywords: adjective, argument structure, kind modifier, modification, nominalization, lexical semantics, logistic regression model, statistical analysis

1   Introduction

Though surprising given standard assumptions about the formal semantics of adjectives, denominal relational adjectives, such as *molecular* and the so-called ethnic adjective (EA) subclass of these, e.g. *French*, have been repeatedly claimed in the syntactic literature to be able to saturate the presumably e-type arguments of the nouns they combine with [16, 6, 10, 1]. On such a view, the compositional contribution of e.g. *French* in (1-a) would be identical to that of the PP in (1-b), saturating the agent argument of *agreement*, though specific semantic analyses for how this is done are not offered in this literature.

(1)   a.   French agreement to participate in the negotiations
       b.   agreement by France to participate in the negotiations

[3], in contrast, focusing specifically on EAs, defend the null hypothesis that EAs are ordinary adjectives that simply modify the noun they combine with, such that the argument-like interpretation arises from the interaction between the semantics of the nominalization and that of the adjective. The goal of this paper is to show that large-scale, statistical corpus data analysis supports the latter analysis.

2   The two competing analyses

We begin by sketching the main features of the argument-saturating analysis and the modifier analyses. For reasons of space, we only discuss [1] to exemplify the former and [3] for the latter, but our comments should generalize to other analyses that preserve the key features of each type of analysis.[3]

2.1   The argument-saturating analysis

[1], working in the framework of Distributed Morphology, treat thematically-used EAs such as those in (1-a)[4] as covert nominals whose nominal source is visible at the level of interpretation. They bear the agent thematic role assigned to them by the (deverbal) noun they modify but lack case. Since every noun needs case, these nouns are deficient and become adjectives in the course of the syntactic derivation. The analysis of a (Greek) example of theirs is shown in (2).

(2)   a.   germaniki epithesi
            German   attack
       b. $[_{DP} [...] [_{FP/AGRP}[_{a(sp)P}[_{a(sp)'}[_{a(sp)^0}$ german$_1$ $[_{a(sp)^0}$-ik]]]]
            $[_{F'}$ F $[_{nP}[_{DP}t_1[_{n'}$ n $[_{vP}$ v $\sqrt{EPITH}]]]]]]$

German- starts out as a DP in the specifier of the noun phrase *epithesi* 'attack', represented in (2-b) via the root $\sqrt{EPITH}$.[5] In this position, *german-* is necessarily assigned the agent theta role by the underlying verb, on analogy to genitive DPs, which are also generated in this position. Since *german-* is not valued for case, it is forced to move up and to adjoin as a head to a(sp), the head of an adjectival projection that generally occupies the specifier position of a functional category between D and N, where it is spelled out as an adjective. On this account, both thematically-used EAs and genitive DPs (e.g. *the Germans' (attack)*) or PPs (e.g. *(the attack) of the Germans*) are base-generated in the same position, hence their relation to the nominal they combine with is the same, namely they saturate the agent argument of the nominal.

One argument put forward for nominal approaches to EAs is the fact that they do not behave like typical adjectives in some respects. For example, EAs cannot be used predicatively ((3-a)), are not gradable ((3-b)) and cannot be coordinated with 'normal' adjectives ((3-c)) (examples from [1]).

(3)   a.   *The intervention in Cyprus was American.
       b.   *the very / more American invasion
       c.   *the immediate / quick / possible and American intervention

---

[3] This section draws heavily on [3]; see that work for further details and discussion.

[4] EAs have been argued to have a second, classificatory use (e.g. *French* classifies wine in *French wine*); see references cited for discussion. [1] treat such adjectives as merely homophonous to thematically-used EAs, whereas e.g. [10] and [3] provide (contrasting) uniform accounts.

[5] Presumably, the nominalizing suffix *-esi* sits in *n*; these details are left out in [1].

However, a serious problem for the nominal account is that the EA does not behave like a typical noun, either. Already [18] noted that EAs are 'anaphoric islands': the alleged nominal underlying the EA does not license anaphora ((4-a)). This fact is also acknowledged by [1], and they provide further examples that show the failure of EAs to bind reflexives ((4-b)), antecede personal pronouns, or control a relative pronoun (see [1] for examples).

(4) a. *The American$_i$ proposal to the UN reveals its$_i$/her$_i$ rigid position.
America$_i$'s proposal to the UN reveals its$_i$/her$_i$ rigid position.
 b. The Albanian destruction (*of itself) grieved the expatriot community.

[1] argue that the status of EAs as anaphoric islands follows from the fact that the underlying noun is morphologically deficient, which results in it becoming an adjective in the course of the derivation. They stipulate that the resulting 'adjective' is deprived of typical nominal anaphoric properties, and that anaphoric rules are sensitive to surface structure configurations only, even though for argument-saturation purposes the nominal nature of EAs is still visible.

We consider this an inelegant solution at best. The facts in (3) do not force the abandonment of an adjectival analysis of EAs in favor of a nominal one, given that many indisputable adjectives (e.g. *alleged*, *main*, *other*) display similar properties (see also [11]); and since there is ample evidence that non-quantificational saturators of e-type arguments quite systematically license discourse anaphora to token individuals, [1]'s analysis effectively renders EAs an anomaly. There are therefore good reasons to explore alternative analyses.

2.2   The modifier analysis

Building on [17] and others, [3] posit that nouns denote descriptions of kinds of individuals. EAs modify these descriptions, introducing a contextually-valued relation (R in (5-b)) between the kind described by the noun and the country, ethnicity, etc.[6] associated with the EA ($\mathbf{R}$ is [8]'s realization relation). Number turns the resulting property of kinds into a property of token entities ((5-d,e)).

(5) a. *agreement*: $\lambda x_k[\mathbf{agreement}(x_k)]$
 b. *French*: $\lambda P_k \lambda x_k[P_k(x_k) \wedge R(x_k, \mathbf{France})]$
 c. [$_{NP}$ *French agreement*]: $\lambda x_k[\mathbf{agreement}(x_k) \wedge R(x_k, \mathbf{France})]$
 d. Num^0: $\lambda P_k \lambda y_o \exists x_k[P_k(x_k) \wedge \mathbf{R}(y_o, x_k)]$
 e. [$_{NumP}$ *French agreement*]: $\lambda y_o \exists x_k[\mathbf{agreement}(x_k) \wedge R(x_k, \mathbf{France}) \wedge \mathbf{R}(y_o, x_k)]$

The relation R is the source of the argument-saturating effect, as nothing in principle prevents it from corresponding to a thematic relation when the modified nominal describes a kind of eventuality. Since EAs do not denote entities, the

---

[6] For brevity, hereafter we refer to this entity simply as the *country*, since our study involved only names of countries.

analysis predicts their general failure to introduce discourse referents, and thus their failure to license discourse anaphora.

Essential to understanding the relative distribution of EAs vs. PPs is the observation that the value of R is restricted, as seen in the strong preference of a PP to describe a visit to Canada out of context, as in (6-a), and in the oddness of an EA in the same context ((6-b)).

(6)   a.   Yeltsin met the prospective Democratic presidential candidate Bill Clinton on June 18. His itinerary also included **an official visit to Canada**. *(BNC)*

       b.   ... His itinerary also included ??**an official Canadian visit**.

To account for this and similar restrictions on the classificatory use of EAs, [3] argue that in the default case, R is the **Origin** relation defined in (7).

(7)   **Origin**(x, y) iff x comes into existence within the spatial domain of y.

Crucially, they posit that not only kinds of concrete objects (e.g. *French bread*) but also kinds of eventualities participate in this relation, by suggesting that the agent(-like) participant in a kind of eventuality can be considered its origin. **Origin** is a *default* value for R because other interpretations are possible when prior discourse makes it clear what that specific relation is between the country and the referent of the head noun (see [3] for examples). Conceiving of the agent relation as a subcase of a default **Origin** value for R allows [3] to provide a unified account of both the basic semantics for the thematic and classificatory uses of EAs and the restrictions on the specific interpretations under these uses.

In our empirical study, we looked at three contrasting predictions of the two analyses that could be translated into features that could be automatically extracted from a corpus containing only morphosyntactic annotation.

2.3   Predictions

Prediction 1. The argument-saturating analysis predicts the distribution of EAs vs. PPs to be roughly the same with event nominals, all other things being equal, given that both are treated as nominals that saturate an argument of the noun they modify. In contrast, the modifier analysis predicts event nominals to combine less often with EAs because when the event nominal has argument structure that must be saturated (i.e. is complex in the sense of [12]), the EA will not be able to do the job, and thus a PP will be required.

Prediction 2. Since, in nondefault cases, the modifier analysis relies on context to supply the identity of the relation between the referent of the head nominal and the country, this analysis predicts the distribution of EAs to be more restricted than that of PPs, as the latter make the relation explicit via the preposition, not relying on context in the same way (recall (6)). Specifically, EAs should occur only when the relevant relation (R in (5-b)) is default or entailed by prior context. In contrast, the argument-saturating analysis predicts, all things being equal, no sensitivity to context in the distribution of EAs vs. PPs.

Prediction 3. A third difference in prediction is specific to analyses such as [3]'s that treat the EA as a modifier of kind descriptions as opposed to token descriptions. A modifier of kind descriptions produces a description of subkinds of the modified kind description. For example, *French bread* and *Italian bread* describe subkinds of bread. We assume that there must be nontrivial criteria that motivate the use of a subkind: a certain number of recognizable instances, an "act of baptism", a recognizable property that characterizes the subkind in contrast to other subkinds, etc. We therefore expect the use of EAs to be concentrated in a comparatively smaller number of nouns than the PPs, i.e. those that meet these criteria, rather than being thinly and evenly distributed across all nouns, as would be expected if the EA could denote a modifier of any token individual description. Conversely, though we see no reason in principle for PPs whose DP complements denote token individuals to be prohibited from serving as a modifier of kind descriptions, we also see no reason for them not to be used as complements or modifiers to descriptions of token entities. The argument-saturating analysis, again all things being equal, does not predict this asymmetric distribution, as it does not provide any basis to distinguish EAs and PPs in terms of the sorts of descriptions they can combine with.

In a previous study on the British National Corpus that did not employ a statistical model [5], we found that this prediction was in fact borne out and that the effect was even more pronounced with EAs with a low frequency and with event nominals. From this we concluded that use of the EA positively correlates with concept stability, i.e. the degree to which the full noun phrase describes a well-established subkind of (abstract or concrete) entities. We posited 1) that stable subkinds describable with EAs are unlikely to be formed for events (e.g. we do not classify agreements according to who makes them) and 2) that fewer stable concepts are formed for those countries we talk about less. However, as will be discussed in Section 4, with the statistical model we obtain different results, so this is a parameter that needs further exploration.

These different predictions grounded our decisions about which features in the corpus to include in the statistical analysis. For Prediction 1, the feature was whether the noun was an event nominal or not. For Prediction 2, since our corpus lacks any semantic annotation, we approximated prior contextual entailment of the value for R with features that correlate with prior mention of the relation: the definiteness of the DP containing the target EA/PP and prior mention of the EA, the head noun, and the corresponding country noun. We also could not test Prediction 3 directly because it is sensitive to the number of *types* of lemmata, and our model operates on the *token* level. To approximate type frequency, we chose the frequency of the head noun as a factor, on the hypothesis that well-established concept descriptions will tend to be formed with nouns of a higher frequency, and thus that EAs will occur more often with these nouns than with low frequency nouns, once possible collocational relations between the EA and head noun as well as effects due to variation in the overall token distribution of the EA vs. the corresponding country noun are controlled for. We will see, however, that this hypothesis turned out to be incorrect.

3 Method

We conducted a study on the British National Corpus (BNC)[7] in order to determine the factors influencing the choice between an EA and a PP, for a sample of 44 different countries whose adjective (e.g., *French*) and proper noun (*France*) forms occur between 1,000 and 30,000 times in the BNC.[8] We tested the predictions outlined in the previous section by defining *features* that could be automatically extracted from a corpus (by running computer programs on the information contained in the BNC) and tested as *factors* in a statistical model. For instance, to test for definiteness of the NP containing the EA or PP (for Prediction 2), we searched for the words *the*, *this*, *that*, *these*, and *those*, followed by at most 4 (for EAs) or 5 (for PPs) arbitrary words excluding verbs, nouns, prepositions, pronouns, determiners, subordinating conjunctions, and punctuation (but optionally allowing for a comma after each adjective) preceding the relevant EA or PP. This type of approach is *noisy*, that is, the information thus gathered is just an approximation of the real syntactico-semantic information that we want to model. However, it has the advantadge that it can be applied on a large scale and that it can be refined and extended with very little effort, so that very different types of information can be explicitly coded and tested.

Our model contains information for the 74,094 occurrences of the relevant adjectives and prepositional phrases (*target expressions* from now on) found in the BNC. These data were analyzed with a logistic regression model [13], which predicts the probability of an adjective realization based on the specified factors and their interactions. Logistic regression has recently become a popular approach for the analysis of similar binary choice problems in quantitative linguistic studies, such as the English dative alternation [7].[9] For model fitting and analysis, we use the R package *rms* [14].

Our best model used 9 factors, which are listed below and grouped according to the theoretical predictions they are connected to. The factor labels shown here will also be used in the presentation and discussion of the results in Section 4.

Prediction 1: tco1: the semantic sort of the head noun, according to the Top Concept Ontology resource (TCO, [2]). This resource restructures the noun hierarchy in WordNet 1.6[10] into a coarse-grained ontology. We only use the highest level of the TCO concept hierarchy, which divides the nominal domain into, roughly, object, event, and abstract nouns.

Prediction 2: definite: the definiteness of the NP containing the target expression, defined as explained above; **recent-mod**: distance, in number of words, to the last mention of the target expression in the same discourse,

---

rescaled as a "decaying activation" $10/(9 + distance)$, or 0 if no previous mention is found; **last-mod-equiv**: whether the previous mention is of the same form (adjective or noun) as the target expression; **recent-head**: same as *recent-mod*, but for the head noun and rescaled as $1/distance$.[11]

Prediction 3: collocAN: the collocational strength between head noun and EA, measured by a conservative estimate of pointwise mutual information[12]; **nhead**: total frequency of the head noun (log-transformed); **ntotal**: total frequency of the target expression (both adjective and noun form, log-transformed); **log-odds-ea-country**: ratio between the frequencies of adjective and noun forms of the target expression in the corpus (log-transformed).

4 Results

From a logistic regression analysis, two basic insights can be gained: (i) which factors or combinations of factors ("interactions") play a significant role in the choice between EA and PP; and (ii) for each significant factor, to what extent and in which manner it increases or decreases the likelihood of an adjective realization (the "partial effect" of the factor).

Prediction	Factor	Chi-Square	d.f.	P
1	tco1	2507.40	2	<.0001
2	definite	42.19	1	<.0001
	recent-mod	658.74	1	<.0001
	last-mod-equiv	46.39	1	<.0001
	recent-head	20.76	1	<.0001
3	collocAN	14.30	1	0.0002
	nhead	648.24	1	<.0001
	ntotal	8.37	1	0.0038
	log-odds-ea-country	1230.93	1	<.0001
	TOTAL	5598.08	10	<.0001

Table 1. Logistic Regression Model: Results of an ANOVA test on the model (Wald Statistics, response: EA). Data distribution: EAs: 51,946 datapoints (70%), nouns (PPs): 22,148 datapoints. See Section 3 for the interpretation of each factor.

Following [13, Ch. 10], we test the significance of factors and interactions by analysis of variance based on the asymptotic standard errors of coefficient estimates (so-called Wald statistics). Table 1 shows highly significant effects for all 9 predictive factors included in the model, lending initial support to the modifier analysis. There is no clear evidence for an interaction between the factors: most interaction terms are not significant or only weakly significant (not

---

[11] The different scaling formulae for *recent-mod* and *recent-head* were found by manual experimentation and resulted in a better fit of the logistic regression model.

[12] This measure compares the observed number of co-occurrences of two words against their expected number of co-occurrences assuming independence; see [15] and [9, p. 86ff.] for details.

shown in the table). Considering the large sample size, we feel that inclusion of such interactions in the model is not justified at this point.

Figure 1 displays a graphical representation of the partial effect of each factor on the likelihood of an adjective realization. In the log odds scale, a value of 0 corresponds to equal likelihood of adjective and noun; positive values indicate that an EA is more probable than a PP. The baseline adjective likelihood of 70% corresponds to a log odds slightly below 1. For example, the middle left panel shows that speakers are more likely to use an adjective with low-frequency head nouns. The same holds if there was a prior mention of the target expression in the same discourse (top center panel). Prior mention of the head noun has an opposite effect (top left panel), but the shaded confidence band around the line indicates considerable uncertainty.
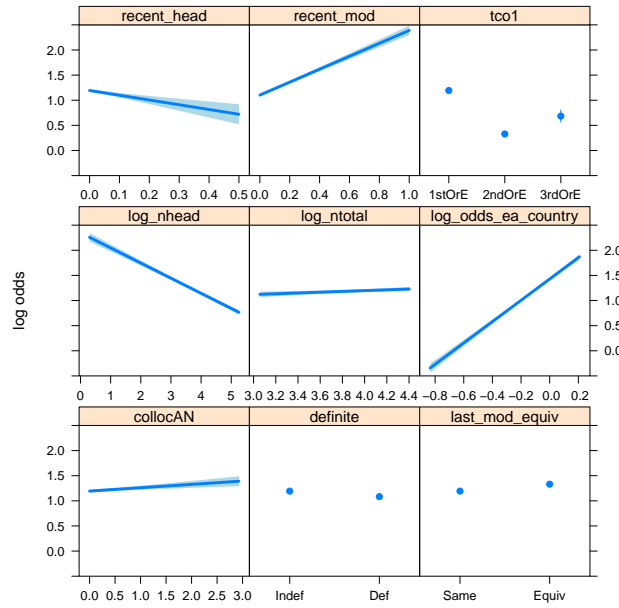

Fig. 1. Logistic Regression Model: graphical representation of the effect of each factor.

Prediction 1 is borne out, as event-denoting nouns strongly disprefer EAs: In the top right panel of Figure 1, adjective likelihood is considerably lower for 2nd Order Entities (the TCO equivalent of event-related nouns) than for 1st Order Entities (object-denoting nouns in the TCO). This is also the most significant effect in Table 1. The results also support Prediction 2, as (a) NPs with definite determiners (bottom center panel in Figure 1) have a slight preference for PPs, (b) a recent mention of the target expression (modifier) favors the use of an EA (top center panel), and (c) a prior mention of the head noun is associated with a PP rather than EA realization (top left panel).[13]

---

[13] It has to be noted that (a) and (c) are relatively small effects despite their high significance, so they do not provide strong evidence either for either analysis.

However, the results do not support Prediction 3, as (a) the collocational strength between the EA and head noun has only a very small effect (bottom left panel), (b) frequent head nouns typically prefer the PP realization rather than the EA (contrary to what we expect based on a modifier analysis; middle left panel), and (c) the overall frequency of the target expression has virtually no effect (whereas we expected more familiar countries to combine more readily with EAs than infrequent countries; middle center panel). However, manual inspection of the corpus data revealed that many of the apparent counterexamples to Prediction 3 involve descriptions of unique individuals (e.g., *Gulf of Mexico*), which have been predicted to resist EAs [3]. Further examination of these data may thus lead to better results. Finally, lexical effects also seem to play an important role, as the overall corpus ratio of adjective/noun expressions has a very large effect on the EA likelihood (middle right panel).

The goodness-of-fit of the logistic regression model is not quite satisfactory yet, with a Nagelkerke R^2 of 11.5% and c index of 67.5% indicating low discriminative power [13, p. 247].[14] As explained in the previous section, all the data used in this study were automatically obtained, a method that, for semantic analysis in particular, adds noise to the data set insofar as the factors are sometimes imperfectly correlated with meaning. Moreover, we are testing the effect of only 9 factors for more than 74,000 tokens. As the choice between EA and PP is a multi-faceted and intricate problem, we expect that adding more factors will improve the predictive power of the model. In particular, we plan to add factors such as posterior mention of the EA/country noun (as discourse topichood of the country may be a factor) and the presence of other modifiers in the target NP (since syntax allows only one EA per noun phrase).

5   Conclusions

We presented statistical support for an account of ethnic adjectives as modifiers, as opposed to saturators. We tested three predictions of the modifier analysis on automatically extracted data for over 70,000 phrases and showed support for two of them; however, the statistical model is still not predictive enough. We have discussed ways in which the model can be enhanced in the future. Given that automatically extracted information is noisy, another enhancement we are planning is to build a supplementary statistical model on a much smaller, manually annotated set of data points. The annotation will include information that cannot be automatically extracted and that has been found to be relevant in the analysis of the results, e.g., whether the NP denotes a unique entity.

---

[14] The c index shows how well the model can discriminate between an example of an EA realization and an example of a PP realization, with a value of 0.5 corresponding to random guesses. It has been shown that c is identical to another popular evaluation measure, the area under ROC curve.

References

1. Alexiadou, A., Stavrou, M.: Ethnic adjectivs as pseudo-adjectives: A case study in syntax-morphology interaction and the structure of DP. Studia Linguistica 65, 1–30 (2011)
2. Álvez, J., Atserias, J., Carrera, J., Climent, S., Laparra, E., Oliver, A., Rigau, G.: Complete and consistent annotation of wordnet using the top concept ontology. In: 6th Language Resources and Evaluation Conference (LREC 2008). Marrakech, Morocco (2008)
3. Arsenijević, B., Boleda, G., Gehrke, B., McNally, L.: Ethnic adjectives are proper adjectives. In: Proceedings of CLS 46 (to appear)
4. Baayen, R.H.: Analyzing Linguistic Data. A Practical Introduction to Statistics Using R. Cambridge University Press, Cambridge (2008)
5. Berndt, D., Boleda, G., Gehrke, B., McNally, L.: Semantic factors in the choice between ethnic adjectives and PP counterparts: Quantitative evidence (2011), paper presented at QITL-4, Berlin, March 2011
6. Bosque, I., Picallo, C.: Postnominal adjectives in Spanish DPs. Journal of Linguistics 32, 349–386 (1996)
7. Bresnan, J., Cueni, A., Nikitina, T., Baayen, R.H.: Predicting the dative alternation. In: Bouma, G., Krämer, I., Zwarts, J. (eds.) Cognitive Foundations of Interpretation, pp. 69–94. Royal Netherlands Academy of Science, Amsterdam (2007)
8. Carlson, G.N.: Reference to Kinds in English. Ph.D. thesis, University of Massachusetts at Amherst (1977)
9. Evert, S.: The Statistics of Word Cooccurrences: Word Pairs and Collocations. Dissertation, Institut für maschinelle Sprachverarbeitung, University of Stuttgart (2004), published in 2005, URN urn:nbn:de:bsz:93-opus-23714.
10. Fábregas, A.: The internal syntactic structure of relational adjectives. Probus 19.1, 135–170 (2007)
11. Gehrke, B., McNally, L.: Frequency adjectives and assertions about event types. In: Proceedings of SALT 19. pp. 180–197. CLC Publications, Ithaca, NY (2011)
12. Grimshaw, J.: Argument Structure. MIT Press, Cambridge, MA (1990)
13. Harrell, F.: Regression Modeling Strategies. Springer, Berlin (2001)
14. Harrell, F.: rms: Regression Modeling Strategies (2011), `http://CRAN.R-project.org/package=rms`, r package version 3.3-2
15. Johnson, M.: Trading recall for precision with confidence sets (2001), unpublished technical report
16. Kayne, R.: On certain differences between French and English. Linguistic Inquiry 12, 349–371 (1981)
17. McNally, L., Boleda, G.: Relational adjectives as properties of kinds. In: Bonami, O., Hofherr, P.C. (eds.) Empirical Issues in Syntax and Semantics, vol. 5, pp. 179–196. http://www.cssp.cnrs.fr/eiss5 (2004)
18. Postal, P.: Anaphoric islands. In: Binnick, R.I. (ed.) Proceedings of the Fifth Regional Meeting of the Chicago Linguistics Society, pp. 205–239. University of Chicago, Department of Linguistics, Chicago (1969)