Greenberg Revisited: Diachronic development of article systems & the structure of DP.

Caitlin Keenan – Harvard University

This paper examines the diachronic development of article systems and the implications of that development for the synchronic analysis of DP. Historical data from several Romance and Germanic languages are introduced which lead to an analysis of article development based on Greenberg's (1978) 'cycle of definiteness'. I show that in all of the languages under investigation, definite and indefinite articles emerge and develop in a consistent, predictable order based on diminishing discourse relevance. I then examine the ramifications of this analysis for our understanding of D-systems cross-linguistically. I argue that the data provide evidence for a unified theory of article development, and against accounts which treat the definite and indefinite articles as unrelated functional items which only accidentally exist in complementary distribution. I analyze the modern (in)definite articles as stronger and weaker versions of a 'discourse reference-marking' article, which resides in the D head.

Data. Historical grammars of French, Spanish, German, and English indicate that the following generalizations hold true for the emergence of article systems in each of those languages:

- emergence of definite articles coincides with breakdown of synthetic systems in Late Latin/Common Germanic
- definite articles emerge significantly[1] before indefinite articles
- indefinite articles are first used in <u>specific indefinite</u> contexts only
- indefinite articles are not used consistently in nonspecific contexts until close to the modern era

Analysis of representative texts spanning 1200 years of linguistic history support these generalizations[2]. The following pattern for English is representative of all four languages:

	Definite	Indefinite (Spec)	Indefinite (Non-spec)
9th-10th c. CE (Alfred translations)	frequently marked (discourse anaphors)	very rarely marked (1/26)	unmarked
11th-13th c. CE (Anglo-Saxon chronicle)	marked (except generics, some autodefinites)	infrequently marked (3/8)	unmarked
14th-15th c. CE (Canterbury tales)	marked (some generics excepted)	usually marked (11/12)	infrequently marked
16th c. CE	marked	marked	frequently marked
Present Day	marked	marked	marked

(15) Ða cusen hi **an clerc** - Whillelm of Curboil was gehaten, he was <u>canonie</u> of **an mynstre** Cicc hatte - brohten him toforen SE KYNG... On ta ilca tyma com **an legat** of Rome - Henri was gehaten, he was <u>ABBOT</u> of Sancte Iohannis mynstre of Anieli...

*Then they chose **a cleric** called William of Curbeil: h was <u>a canon</u> from **a monastery** called St Osyth's. They presented him to the king... At the same time came **a certain legate** from Rome, who was called Henry. He was abbot of the abbey of St Jean d'Angely...*   PETERBOROUGH CHRONICLE (1123 CE)

The indefinite article use in this text – with three specific NPs marked with the article and the single non-specific NP unmarked – is typical of all four languages in the Middle period.

The (Neo)Greenbergian Model. In the process of grammaticalization described in Greenberg (1978), demonstratives consistently develop into definite articles, general articles, and eventually gender markers on nouns. Greenberg's data is drawn from a range of African and American languages with single-article systems. Previous attempts to extend this proposal to the analysis of Romance and Germanic articles have focused only on the development of the definite article (Harris 1980). The present paper will argue that Greenberg's cycle of article development can only be appropriately

---

[1] At least 200 years elapse between the first consistent use of the definite article and the first consistent use of the indefinite article in each language.

[2] Texts analysed (in whole or part) in the course of this research to date include: *OE* – Bede's account of the poet Caedmon & *The Voyages of Ohthere and Wulfstan* (both tr. by King Alfred, 9th c. CE); *Early ME* – excerpts from the *Anglo-Saxon Chronicle*, esp. the Peterborough Chronicle (12th c. CE); *Late ME* – excerpts from *The Canterbury Tales* (Chaucer, turn of the 15th c. CE); *OF* – *La Cantilène de Ste Eulalie* (ca 878 CE); *MF* – *Le Chanson de Roland* (12th c. CE); *OHG* – excerpts from *Hildebrandslied* (9th c. CE); *MHG* – poems of von der Vogelweide (12th-13th c. CE); *Early NHG* – Luther's Bible (1545 CE); *OS* – *Glosas Emilienenses* (10th c. CE); *Early MS* – excerpts from *El Canto Mio Cid* (12th c. CE); *Late MS* – excerpts from *Don Quixote* (de Cervantes, 1605 CE).

extended to languages with rich article systems if the *entire* article system of each language is taken into consideration. Greenberg's languages developed a discourse definite article (Stage 0-1), which spread progressively to non-anaphoric definite contexts (late Stage 1), specific contexts (Stage 2), and finally to all nouns regardless of context (Stage 3). This is precisely the semantic order in which articles in the Romance and Germanic languages develop; whereas Greenberg's languages extend a single article through the entire paradigm, Romance and Germanic develop a new article at Stage 2:
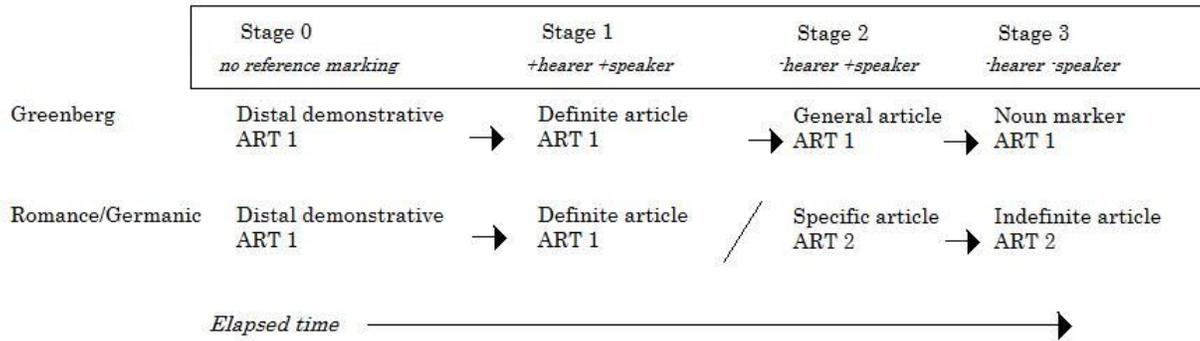
(Fig.1) Stages in the development of reference-marking articles

Once a language has begun – for whatever reason – to mark discourse reference overtly in the functional structure, its use of articles in this function will gradually spread in a predictable manner to a broader (and progressively less referentially 'strong') range of noun phrases. I propose that whenever an article-developing language broadens its article use to the point where a new 'stage' is to be overtly marked, the language has two options: it can either continue to broaden the meaning of the original article (at the expense of clarity) or introduce a new article to mark the new meaning (at the expense of simplicity). This is the age-old choice between economy and explicitness. Whereas Greenberg's languages favour economy and developed a single article which eventually becomes a ubiquitous nominal (gender) marker, English, German, Spanish, and French compromise between economy and explicitness. They develop a Stage-I definite article and a separate Stage-II specific article (preserving a semantic distinction), but subsequently extend the latter article to Stage-III, giving the ambiguous specific/non-specific indefinite article found in all the modern languages.

Syntactic Analysis. The data presented in this paper show that definite and indefinite articles are crucially related to each other both semantically and structurally. The closeness of this relationship has consequences for the synchronic analysis of articles in the Determiner Phrase. Certain recent proposals (Lyons 1999, Wood 2003) have argued that the apparent complementary distribution of definite and indefinite articles in a language like English is incidental, and that in fact these two article types are exponents of separate functional projections: for instance, the definite article may be head D, while the indefinite article resides in a lower functional projection, such as NumP. The intimate relationship between definite and indefinite articles uncovered by the historic data in this paper argues strongly against such proposals. The definite and indefinite articles exist in complementary distribution, have related discourse reference functions, and emerge in a principled way relative to one another. Based on these facts, I propose that articles of different strengths are simply complementary exponents of the same basic feature bundle, [Discourse Reference]. As discourse reference marking grammaticalizes, overt D is required in a wider variety of circumstances, and articles are co-opted from elsewhere in the lexical system to fulfill this function. Although indefinite articles may merge into the structure below DP, both diachronic and synchronic evidence point to a principled relationship between the D head and the article system as a whole. I conclude the paper with a discussion of synchronic data from several modern Chinese languages which support the cross-linguistic applicability of this analysis.

Select References. **(1)** Greenberg, J. (1978). "How does a language acquire gender markers?" In *Universals of Human Language Volume 3: Word Structure,* J. Greenberg (ed). (49-81). **(2)** Harris, M.(1980). "The marking of definiteness: a diachronic perspective". In *Papers from the Fourth International Conference on Historical Linguistics*, E. Traugott, R. Labrum and S. Shepherd (eds).Amsterdam: John Benjamins B.V. **(3)** Huang, Li, and Li. (2009). *The Syntax of Chinese*. Cambridge: Cambridge University Press. **(4)** Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press. **(5)** Wood, J. (2003). *Definiteness and Number: Determiner Phrase and Number Phrase in the History of English*. Doctoral dissertation. Arizona State University.