

Studies in the Linguistic Sciences

VOLUME 32, NUMBER 1
(SPRING 2002)

PAPERS IN GENERAL LINGUISTICS



DEPARTMENT OF LINGUISTICS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

STUDIES IN THE LINGUISTIC SCIENCES
(ISSN 0049-2388)

PUBLICATION OF THE DEPARTMENT OF LINGUISTICS
WITH SUPPORT OF THE HUMANITIES COUNCIL IN THE
COLLEGE OF LIBERAL ARTS AND SCIENCES OF THE
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

MANAGING EDITOR: Peter Lasersohn
EDITORIAL ASSISTANTS: Aimee Johansen, Angela Nollett

EDITORIAL BOARD: Elabbas Benmamoun, Rakesh Bhatt, Eyamba G. Bokamba, Chin-Chuan Cheng, Jennifer S. Cole, Adele Goldberg, Georgia M. Green, Hans Henrich Hock, José I. Hualde, Braj B. Kachru, Yamuna Kachru, Chin-W. Kim, Charles W. Kisseberth, Peter Lasersohn, Molly Mack, Howard Maclay, Jerry L. Morgan, Rajeshwari Pandharipande, Chilin Shih, Daniel Silverman, Richard Sproat, James H. Yoon, and Ladislav Zgusta.

AIM: *SLS* is intended as a forum for the presentation of the latest original research by the faculty and students of the Department of Linguistics, University of Illinois, Urbana-Champaign. Scholars outside the Department and from other institutions are also cordially invited to submit original linguistic research for consideration. In all cases, articles submitted for publication will be reviewed by a panel of at least two experts in the appropriate field to determine suitability for publication. Copyright remains with the individual authors. Authors of articles will receive one copy of the particular issue and 10 offprints of their individual contributions.

SLS appears twice a year, and one issue is traditionally devoted to restricted, specialized topics. A complete list of available back issues is on the last page of this issue.

BOOKS FOR REVIEW: Review copies of books may be sent to:
Review Editor, *Studies in the Linguistic Sciences*
Department of Linguistics, 4080 For. Lang. Bldg.,
University of Illinois
707 S. Mathews,
Urbana, IL 61801, USA

SUBSCRIPTION: Requests for subscriptions should be addressed to *SLS* Subscriptions, Department of Linguistics, 4080 Foreign Languages Building, University of Illinois, 707 South Mathews, Urbana, Illinois 61801.

Price: \$15.00 per single issue.

e-mail address: slingsci@uiuc.edu

Telephone: (217) 333-3563

Fax: (217) 333-3466

Homepage: <http://www.linguistics.uiuc.edu/sls.html>

STUDIES IN THE LINGUISTIC SCIENCES

Papers

in

General Linguistics

EDITED BY

Peter Lasersohn


EDITORIAL ASSISTANTS

Aimee Johansen

Angela Nollett

**VOLUME 32, NUMBER 1
(SPRING 2002)**

DEPARTMENT OF LINGUISTICS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN



Digitized by the Internet Archive
in 2011 with funding from
University of Illinois Urbana-Champaign

TABLE OF CONTENTS

Zsuzsanna Fagyal, Noël Nguyen, and Philippe Boula de Mareüil: From <i>dilation</i> to coarticulation: Is there vowel harmony in French?	1
Liwei Gao: On the nativization of English in China	23
José Ignacio Hualde, Pello Mugarza, and Koldo Zuazo: The accentual system of Mallabia Basque	39
Yong-hun Lee: Resolution algorithms for the Korean reflexive <i>cakicasin</i> : A Categorical Grammar approach	57
Theeraporn Ratitamkul: An alternative analysis to the temporal adverbial <i>when</i> -clause	79

REVIEWS

José Ignacio Hualde: Review of R. M. W. Dixon & Alexandra Y. Aikhenvald: <i>Word: A Cross-Linguistic Typology</i>	91
Cedric Boeckx: Review of Thomas Ernst: <i>The Syntax of Adjuncts</i>	95
José Ignacio Hualde: Review of Robert Stockwell and Donka Minkova: <i>English words: History and Structure</i>	103

**FROM DILATION TO COARTICULATION:
IS THERE VOWEL HARMONY IN FRENCH?**

Zsuzsanna Fagyal*, Noël Nguyen†, and Philippe Boula de Mareuil‡

**University of Illinois at Urbana-Champaign*

zsfagyal@uiuc.edu

†*Laboratoire Parole et Langage, CNRS & Université de Provence,*

Aix-en-Provence, France

‡*LIMSI, CNRS, Orsay*

This paper presents the preliminary results of an acoustic study, and a review of previous work on vowel harmony in French. It shows that harmony, initially regarded as regular sound change, is considered an optional constraint on the distribution of mid vowels. Acoustic evidence of anticipatory assimilation of pretonic mid vowels to tonic high and low vowels is shown in three speakers' readings of disyllabic words in two dialects. It is argued that vowel-to-vowel assimilation, referred to as 'vowel harmony', does exist in French, and it is likely to extend beyond the morphological contexts in which it was previously thought to operate.

1.0 Introduction

Whether vowel harmony (henceforth, VH) exists in French might seem an incongruous question, since French is not known as one of the classic harmonic systems requiring all vowels in a prosodic domain, typically the word, to share ('harmonize' in) one or several features. While the phonological systems of Turkish and Hungarian, for instance, represent textbook examples of rounding and front/back harmonies, and many African languages, e.g., Igbo, are known for their use of [ATR] harmony, no dialect of French has ever been described as making an extensive use of VH. Yet, historical and contemporary descriptions routinely analyze cases of word-level vowel-to-vowel assimilation in such terms. As a type of vowel assimilation process, VH in these studies can be characterized as a 'form of qualitative articulatory adaptation between vowels with regard to one or more features' (Bussmann 1996:518). While previous studies of the French sound system generally agree on the existence of VH in such a broad sense, they often differ in their analyses of actual data.

VH in French is commonly described as a word-level anticipatory process affecting non-final mid vowels in open syllables that assimilate in height to the vowel of the final, tonic syllable.¹ When the vowel of the tonic syllable is a non-

¹ The French vowel system has four degrees of opening, i.e., mid-vowels have two allophones. These are referred to as *mi-ouvertes* 'half-open' ([e]-[œ]-[ɔ]), and *mi-fermées* 'half-closed' ([e]-[ø]-[o]) in French (see Tranel 1987). Conceptualized as degrees of vowel 'height', these terms are commonly translated as 'low-mid' and 'high-mid' in English (see Walker 2001).

low vowel, such as [e] in examples (1a) and (2a), the pretonic mid vowel should be high-mid, such as [e] in (1a), and [ø] in (2a). Conversely, when the final vowel is low, like [a] in (1b), and [ã] in (2b), the pretonic mid vowel becomes low-mid, resulting in [ɛ] and [æ] in (1b) and (2b) (Walker 2001:54):

- (1a) *aimer* [ɛme] 'to love'
 (1b) *aimable* [ɛmabl] 'lovely'
 (2a) *aveugler* [avøgle] 'to blind'
 (2b) *aveuglant* [avæglã] 'blinding'

VH is also thought to operate in morphologically derived contexts, where the mid vowel of the base may alternate depending on the height of the following suffix vowel. For instance in (3), /o/ is realized as low-mid in the derived form under the influence of the following low-mid vowel [ɛ] (Walker 2001:55):

- (3) *gros* [gʁo] 'huge', adj. masc. > *grossesse* [gʁɔsɛs] 'pregnancy'

Notice, however, that the outcome in (3) conflicts with predictions that can be made in terms of 'open/closed syllable adjustment' (*la loi de position*). According to this constraint — subject to many exceptions in different dialects — the degree of opening of a mid vowel is determined by syllable structure: high-mid vowels (*voyelles mi-fermées*) appear in open syllables, while low-mid vowels (*voyelles mi-ouvertes*) occur in closed syllables. Therefore, the outcome of the derivation in (3) could also be as shown in (4), i.e., with /o/ realized as high-mid:

- (4) *grosse* [gʁo:s] 'huge', adj. fem. > *grossesse* [gʁɔsɛs] 'pregnancy'

Although the input form *grosse* 'huge' in (4), featuring a high-mid vowel in a closed syllable, represents an exception to the open/closed syllable adjustment rule (*la loi de position*), the output is the preferred realization for Tranel (1987:61). For Tranel the pretonic high-mid vowel [o] in (4) is arguably due to some type of faithfulness constraint, as 'the vowel quality found in the base word is preserved in the derived word' (*ibid.*).

As the examples in (3) and (4) illustrate, the distribution of mid vowels in these contexts has been analyzed in terms of several, sometimes conflicting, processes. It is, therefore, an open question whether VH needs to be part of this complex repertoire of generalizations. Crucially, one must ask the question: is there empirical evidence in support of the phenomena illustrated above? Also, if it exists, does vowel assimilation in French represent instances of VH or of other types of assimilatory change? Since the vowel triggering the harmony effect is that of the suffix, i.e., the process is anticipatory, an analysis in terms of umlaut or metaphony² could be equally likely. Furthermore, is VH tied to specific morphological environments or can it also be shown in semantically and morphologically unrelated contexts?

² Some analysts restrict the term 'vowel harmony' to carry-over assimilation, preferring terms such as 'umlaut' or 'affection' for cases of anticipatory assimilation (see Trask 1996:383). This point is discussed in greater detail in the conclusion.

The answer to the first question is straightforward: to the best of our knowledge, there is currently no empirical data on VH in French. Thus, work presented in this paper represents the first such investigation. There is also a general lack of evidence on realizations of mid vowels in different dialects of French despite the fact that mid vowels in some contexts can become actual shibboleth words, allowing a native speaker to tell the approximate geographical origin of another speaker. Front and back rounded mid vowels in words such as *chose* 'thing', *feutre* 'felt', and *veule* 'spineless' are always open ([œ] and [ɔ]) in what is referred to as 'the South', i.e., roughly the Oc dialect area, but they are usually realized as high-mid ([ø] and [o]) in varieties of Île-de-France, i.e., part of the Oil dialect area. A strict North-South division along this pattern, however, does not apply, since vernacular varieties of Belgian French also prefer [œ] and [ɔ] in these contexts (Blampin 1997:169).

In this paper we will examine two types of mid vowel in the pretonic position of existing, disyllabic words in two dialects of French. We will first offer a review of previous studies in order to test some of the earlier claims empirically. We will then present evidence in support of anticipatory vowel assimilation, and suggest relating our findings to a group of phenomena known as metaphony in the history of French and other Romance languages.

2.0 Previous treatments of VH in French

2.1 Dilation, metaphony, and umlaut

Vowel-to-vowel assimilation in French was first analyzed in the *Traité de phonétique* 'Treatise on Phonetics' by Maurice Grammont (1914[1939]) in terms of 'vocalic dilation'. As phonetic dictionaries rightly point out, such terms — now used only sporadically in studies of French phonology (Gadet 1997:74-5) — might be just 'generic label[s] applied in a quite bewildering fashion to a wide range of historical changes' (Trask 1996:221).³ From the Latin *dīlāto* 'to extend', the term 'dilation' refers to the anticipation or persistence of articulatory movements beyond the syllable in which the segment is located. It appears to be a metaphor for assimilation between non-adjacent segments: 'certain qualities of a phoneme spread to another phoneme that is not in immediate contact with the propagating phoneme' (Grammont 1939:251). Cases of vocalic dilation — among them umlaut and metaphony — are viewed as a regular sound change, affecting all morphological classes of a certain phonetic form at once, and with great uniformity: 'Dilation is accomplished with perfect regularity and consistence when all conditions it requires are met' (p. 251). Although the term 'vowel harmony' does not appear in the treatise, the definition provided by Grammont for 'dilation' matches contemporary descriptions of VH in French:

In **Parisian French** [our emphasis], vocalic dilation is regressive; it departs generally from the stressed vowel: the first of the two vowels in two con-

³ The title of the chapter is 'Dilation', which is the preferred spelling of the term throughout the text. One exception is the spelling *dilatation* on p. 251 with respect to assimilation between non-adjacent consonants. Whether this change in spelling signals a meaningful conceptual difference is, however, difficult to tell.

secutive syllables tends to adjust its quality to the second one; this is a dilation in degree of opening, and as such also in point of articulation [the position of the tongue along the front-back axis] (p. 266).

The process is argued to be productive in derivational and inflectional contexts. More typical in the front-mid series, it introduces a systematic [ɛ]/[e] alternation depending on the height of the following vowel, as in (5).

- (5) *plaire* [plɛʁ] 'to please'
plaisir [plɛziʁ] 'pleasure'
plaisant [plɛzɑ̃] 'pleasurable'

Somewhat diminishing his own claim about the categorical nature of this opposition, Grammont confides, however, that the alternation is a rather 'delicate nuance' which might be difficult to perceive, because it affects unaccented vowels 'articulated with little tension' (ibid.). The phenomenon seems less clear for /o/, and is subject to many variations in front vowels, as well.

André Martinet (1945) affirms having found no evidence that the first vowel in a disyllabic word would be affected by the quality of the following vowel. Based on questionnaires circulated in a POW camp for French officers, Martinet declares that in none of the regional pronunciations of French does the assimilation process seem to be typical of 'the majority of the subjects' (p. 142). Therefore, he argues, 'it does not seem to present the universal character many phoneticians, including Mr. Grammont, wish to attribute to it' (ibid.). Notice, however, that the two test cases Martinet presented to his subjects, (6a-b), are examples that Grammont did not include in his discussion of vocalic dilation. The pair in (6a) is not quoted in his treatise, and the example in (6b) is based on the presumed alternation of the mid vowels [œ] and [ø], which is not discussed in Grammont's dilation rule.⁴

- (6a) *été* [ete] 'summer' (il) *était* [ɛte] '(he) was'
 (6b) *déjeuner* [dezøne] 'lunch' *déjeunons* [dezœnɔ̃] '(we) have lunch'

The notion of dilation also appears in *Phonétique historique du français* 'Historical Phonetics of French' by Pierre Fouché (1952). It refers, however, only to historical changes that led, for instance, to the closure and subsequent nasalization of ē to ī in words in Old French (*vēnī* (Lat.) > *vīn* (O.Fr.)).⁵ Synchronic aspects of vowel-to-vowel assimilation are discussed in the *Traité de prononciation française* 'French Pronunciation Treatise' (Fouché 1956) in terms of VH, which is defined similarly to Grammont's conception of dilation. A few differences between the two approaches are, however, worth mentioning.

⁴ We doubt that the alternation in (6b) exists in contemporary French. Based on Fougeron & Smith 1999, the assimilating, final, back, rounded nasal vowel in *déjeunons* is more closed and rounded than Martinet's transcription in [ɔ̃] makes it appear.

⁵ The fronting of Latin ē to ī in Old French (*fēci* (Lat.) > **fīci* > *fīs* (Old Fr.)) is also cited as an example of dilation/VH by Malmberg (1974:178).

Fouché clearly states that VH does not affect the back mid vowel /o/, which remains phonetically low (half-open) no matter what vowel follows in the tonic syllable (7).

(7) *vol* [vɔl] 'flight' > *voler* [vɔle] 'to fly'; *volant* [vɔlɑ̃], 'flying'

This claim is, however, attenuated by pointing to a possible influence of a tonic /l/ on the preceding /o/ in words such as *fossile* 'fossil' and *vomir* 'to vomit'. Closed syllables with a low-mid front vowel before /r/ in pretonic position are also excluded from VH, which means that words such as in (8a-b) will not show the [ɛ]/[e] alternation.

(8a) *(il) perce* [pɛʁs] '(he) drills' > *percer* [pɛʁse] 'to drill'

(8b) *(il) perd* [pɛʁ] '(he) loses' > *perdu* [pɛʁdy] 'lost'

VH is, therefore, restricted to the pretonic front mid vowel [ɛ] in open-syllables, assimilating in frontness/degree of opening to the following high or high-mid front vowel [i], [e], or [y]. Fouché is also the first to consider stylistic factors in VH: 'Vocalic harmonization, I repeat, characterizes conversational speech. The more the pronunciation is careful, the less it is present' (1956:71). This idea will surface in subsequent discussions of VH (Léon 1966:51, Dell 1972:215).

Another original point in Fouché's approach is what appears to be the consideration of prosodic factors. The VH rule is written for the 'stressed vowel [ɛ], spelled *ai*, *aî*, *ei*, *ay*, or *ê* (other than in words with *-ême*, and *e* before *ff*, *ss*, *tt*), becoming *unstressed*, while remaining open or passing to [e] with no alternation in spelling' (p. 70). However, prosodic factors, analyzed for instance in other languages (McCormick 1982; Hualde 1989), are merely an artifact of the formulation of morphological conditions in which VH is thought to operate. When stating that 'vowels can appear in stressed and unstressed syllables within the language' (p. 64), Fouché is, in fact, considering semantically related pairs of words that are in derivational or inflectional morphological relationship with each other. Subdivided into different word classes, monosyllabic base words in these sets contain 'stressed' vowels, which become, by definition, 'unstressed' in the pretonic syllables of disyllabic derived forms (9a-b).

(9a) *fraîche* [fʁɛʃ] 'fresh' > *fraîcheur* [fʁɛʃœʁ] 'freshness'

(9b) *fête* [fɛt] 'party' > *fêter* [fɛte] 'to celebrate'

Thus metrical stress alternation is not a factor, but a correlate of VH. In Fouché's terms, VH is viewed as an analogical process extending a morphophonemic alternation to other roots and base words in the grammar. Such processes, generalizing the results of a previous sound change are well-documented in the history of other languages as well (Hock 1991:187).

All reference to the term 'dilation' disappears in Bertil Malmberg's phonetic treatises. In *La Phonétique* (1966), VH is treated synonymously with metaphony and umlaut, while in *Phonétique Française* (1969), no general definition of VH is provided. It is referred to as an assimilatory process causing the neutralization of the /ɛ/-/e/ distinction in unaccented syllables. Consequently, only examples from the front-mid series are provided. In his *Manuel de phonétique générale* (1974), all

reference to metaphony is omitted, and VH is treated as a case of vocalic dilation similar to umlaut (10).

(10) (*vous laissez* [lese] '(you) leave' / (*nous laissons* [lesɔ̃] '(we) leave')

Yet, the use of four different terms for a possibly unique phenomenon forces Malmberg (1969:178) to comment on the diversity of denominations: 'It is unlikely that all phenomena that received, in Germanic historical phonetics, the name of umlaut, would involve the same phonetic mechanism'. No further comment is, however, provided as to what this phonetic mechanism might be.

2.2 From generative phonology to current approaches

One of the first analyses of VH in French within generative phonology is that of François Dell (1972)⁶, whose VH rule depicts an optional process applying relatively late in the grammar, and affecting only non-rounded front-mid vowels in word-medial position. Contrary to previous analyses, Dell's HARM rule (11) states two important points: the triggering vowel should belong to another morpheme, and the assimilated vowel can be either low or non-low:

[The rule] rewrites [e] to [ɛ] when the following syllable contains a low vowel that does not belong to the same morpheme, and it rewrites [ɛ] to [e] when the following syllable contains a non-low vowel in a separate morpheme (Dell 1972:214).

(11) [+syll, -round, -high, -back] → [α low] / __C₁ + C₀ [+syll, α low]

Similar to Fouché's conception, Dell's VH rule is style-dependent in that in hyper-articulated speech 'it only applies sporadically' (p. 215). Its variable aspect is made clear by the statement that 'a systematic investigation would surely find quite important inter-speaker differences' (ibid.). Based on Dell's examples, most of which also appear in previous treatises, VH does not seem to be blocked by an optional schwa intervening between the assimilated and the triggering vowels. Tri-syllabic words containing a word-medial schwa, for instance *aidez* '(you) will help', *cédez* '(you) will concede', and *lèvez* '(you) will lift', appear among Dell's examples, but due to schwa-lenition applying before the HARM rule, these words are considered disyllabic.

Jean Casagrande's (1984) approach brings further details to generative analyses. It restricts VH to low-mid vowels, insisting that the 'tendency is toward closure' (p. 89), but as a novelty, it also extends it to the rounded mid vowel [œ], as illustrated in the following example:

(12) *beuglement* [bœgləmā] 'lowing' vs. *beugler* [bøgle] 'to low'

Although no examples are supplied for the back mid vowel, for which VH is 'negligible if it exists at all' (ibid.), mid vowels are ranked with respect to their sensitivity to VH. According to the author, VH would be more frequent for [ɛ] than for [œ], and for [œ] than for [ɔ̃]. Casagrande is also the first to explicitly restrict VH to the mid vowel series, arguing that even though a vowel tends to close

⁶ Previous approaches by Morin (1971) and Selkirk (1972) are cited by Dell (1973:215).

under the effect of VH, it can never become [+high] in the process. For instance, the post-tonic vowel in *bêtise* can never become [i] under the effect of VH:

(13) *bête* [bet] 'stupid' > *bêtise* [betiz] but *[bitiz] 'stupidity'

Bernard Tranel's (1987) *The Sounds of French* conceives of VH as one of the many constraints influencing the distribution of mid vowels in non-final syllables. Distancing himself entirely from historical aspects that are not even mentioned, Tranel presents VH as optional even in the front-mid series where other factors, such as spelling, faithfulness to the base, and open/closed-syllable adjustment also play a role (see 1.0). As in all previous approaches, the majority of examples involve disyllabic words. In these, VH is strictly anticipatory and most frequent when the triggering vowel is high or high-mid, and the assimilated vowel is low or low-mid. Cases such as in (14) are typically cast as a high vowel (here [i]) in the tonic syllable triggering the shift of a low vowel (here [ɛ]) to a non-low vowel (here [e]):

(14) *bête* [bet] 'stupid' > *bêtise* [betiz] 'stupidity'

There is also room for variation, since VH is considered optional when the triggering vowel is low or low-mid, and is preceded by a vowel that is high-mid. One such case is illustrated in (15), where the low vowel [ɛ] in the suffix of *théière* may cause the preceding [e] of *thé* to assimilate to [ɛ]:

(15) *thé* [te] 'tea' > *théière* [tejɛʁ] or [tejɛʁ] 'tea kettle'

According to Tranel, the vowels [œ] and [ø] can also undergo VH, but only under the influence of their high-mid counterparts. Two such examples are shown in (16) and (17) where two realizations are possible whether the high-mid vowels in the tonic syllable exert or not their assimilatory influence on the preceding vowels. No examples are quoted for the opening/lowering of high-mid back vowels.

(16) *heureux* [œʁø] or [øʁø] 'happy'

(17) *auto* [otɔ] or [otø] 'car'

Stylistic factors are pointed out, suggesting that VH occurs 'naturally in spontaneous speech, where a relatively rapid and unmonitored delivery is generally maintained' (1987:61). Spelling is a factor on which previous studies relied when subdividing words into different lexical classes, but did not examine. Tranel suggests that some orthographic representations might be directly tied to phonological representations, for instance, 'é favors the pronunciation of [e], whereas the spellings è, ê, ai, and e before two written consonants may favor the pronunciation [ɛ]' (ibid.). Another novelty in this approach is that VH is exemplified mostly in citation forms. Although base words from which possible derivations can stem are sometimes indicated in brackets, VH does not seem to be tied to specific morphological environments; it can apply in a single lexical item, as in (17), as well as a chain of derivation, as in (14). In other words, Tranel's conception seems to be the closest to a view of VH as a non-morphologically driven, general assimilatory process.

With a few modifications, other contemporary studies and treatises subscribe to the interpretation of VH as an anticipatory process affecting primarily front-mid vowels. Landick (1993), the only author taking a lexicological approach to the question, declares that in isolation, e.g., in citation forms in dictionaries, VH is not a significant feature of contemporary French. Walker (2001) presents examples only in derivational and inflectional contexts, while Battye et al. (2000) extend VH to the closing of [ɔ] to [o] under the effect of a following [i] in words, such as *dormir* [dɔʁmiʁ]⁷ and *automobile* [ɔtɔmobil].

In this paper, we intend to test some of these earlier claims empirically. Following up on Grammont's hint on the 'delicate nuance' that VH represents, and Ohala's (1994:4.1) suggestion that VH is a 'fossilized remnant of an earlier phonetic process involving vowel-to-vowel assimilation', we propose an acoustic study of vowel assimilation processes presented above. Similar to our predecessors, we expect an anticipatory assimilation of pretonic mid-vowels to tonic vowels in terms of height, i.e., opening and frontness/backness. As in most previous studies, our corpus will be composed of disyllabic words, and will involve both low and high vowels. While not testing for stylistic variation, we propose to take into account some geographical variation.

3.0 Corpus, method, and expectations

The corpus was composed of 136 pairs of disyllabic nouns, adjectives, and infinitives. The first syllable always contained a mid vowel (henceforth, V1), and was phonemically identical in both words of the pair. The second syllable contained a non-low vowel in one word and a low vowel in the other word of the pair (henceforth, V2 for both). The onset of each syllable of each pair was either a single consonant or a consonant cluster. The complete list of target words grouped in eight sets is shown in Table 5 in the Appendix.

We restrict our analysis to words in which the first syllable was open. Table 1 shows the sets of target word pairs with one pair referring to each set, the underlying representation of V1, the underlying representation of V2, and the total number of word pairs in each set. Vowel quality in V1 is determined following the rule of open-syllable adjustment. Since we are reporting only on words in which the first syllable is open, we expect all V1 to be non-low (high-mid) underlyingly. The labeling of V2 in Table 1 follows the expected alternation of a non-low vowel with a low vowel in Northern Metropolitan French dialects ([e]/[ɛ], [ø]/[œ], and [i]/[a]). This means that the realization of V2 in some of the target words in the Southern speaker's, S3's, readings can be different from this labeling (see 1.0). Words in which the underlying vowel quality of V2 might differ from the Northern pattern are indicated in bold in Table 5 of the Appendix, and will be analyzed among the results.

⁷ Although the phonetic transcription proposed by the authors is [dɔʁmiʁ] (Battye et al. 2000:93), if VH exerts its influence as they suggest it does, the first vowel in *dormir* should be high-mid, as it appears in our transcription.

A total of six speakers, four female and two male, was recorded in two dialects of French. Results of the first acoustic analyses are presented here for three female speakers. Background information about each speaker was obtained via biographic questionnaires administered after the final recording session (Table 2). Based on our interactions with the speakers, we found the vernacular of female speakers S1 and S2 to be typical of the Île-de-France variety of Northern Metropolitan French, while speaker S3's speech represents a Southern variety of French commonly heard in Aix-en-Provence. The speakers' age and educational level were comparable. All lived at their place of permanent residency until at least the end of their secondary education.

Table 1: Type and number of word pairs with a non-final open syllable (119 out of a total of 136 pairs).

set of word pairs	V1	V2	N (=119)
été — éther	e	e / ε	18
prêteuse — prêteur	e	ø / œ	8
dérot — dévotte	e	o / ɔ	6
potée — poterne	o	e / ε	33
poseuse — poseur	o	ø / œ	20
auto — automne	o	o / ɔ	4
épice — épate	e	i / a	14
notice — nota	o	i / a	16

Speakers S1 and S2 were taped in the Phonetics Laboratory of the University of Illinois at Urbana-Champaign. They had been living in the United States for less than three years, using and teaching their native language on a daily basis. Speaker S3 was recorded in the Laboratoire Parole et Langage of the University of Provence in Aix-en-Provence. At both locations, the recordings took place in an anechoic chamber using a high-quality microphone and a DAT recorder (input sampling frequency of 44.1 kHz). Despite our expectations, speaker S3 seemed to have 'style-shifted' during her recording in laboratory condition. At several instances, she seemed to have adopted a reading style close to standard French, dropping some of the expected characteristics of her Southern Metropolitan accent. The possible implications of this reading style are discussed among the results.

The target words were presented to the speakers on slides shown on a computer screen. Each word appeared twice on each slide, and was included in a carrier sentence shown in (18).

(18) *Il retape _____ parfois; _____.*

'He retypes (target word) sometimes; (target word).'

The sentence refers to a man typing words on a keyboard. The first instance of the target word appears between two symmetrical phonetic sequences (/ap/ and /pa/), while the second repetition represents a full intonation phrase. The bilabial stops

to the left and right of the utterance-medial target word were chosen for their minimal influence on the target vowels. The utterance-final target words were intended as stimuli for future perceptual experiments, but results of acoustic analyses of their vowels are also incorporated in this study.

Each speaker was recorded twice in two separate sessions, with at least a week interval between the sessions. Target sentences were presented in varying random orders. The speakers were asked to read each sentence at a normal rate, with no special emphasis on any word. They were instructed not to pause between words in the first phrase (up to *parfois* 'sometimes'). The utterance-final rendition of the target word had to be as similar to the first one as possible, exhibiting the same speech rate, rhythm, and intonation. Although there were no distractor sentences, breaks were incorporated in each reading after thirty slides each time. During these breaks, special effort was made to distract the speaker from his or her reading task. Before the recording started, written instructions to speakers were also explained verbally, and one training sentence was provided to check whether the speaker understood the instructions. The data, four repetitions of each word for each speaker, were then transferred onto the computer with a sampling rate of 22kHz at 16 bits for further computer processing.

Table 2: Speaker characteristics.

	S1	S2	S3
sex	F	F	F
age (years)	28	31	34
birth place	Fort-de-France	Bretagne	Salon-de-Provence
length of years at birth place	0.4	8	20
permanent residency	Paris intra muros	greater Paris	Aix-en-Provence
length of years at residency	22	20	14

The target vowels were manually segmented using the waveform and wide-band spectrograms obtained in the Praat 4.0 and the Entropic ESPS/Xwaves speech analysis programs at the University of Illinois and the University of Provence, respectively. The onset of target vowels preceded by a stop consonant was set to the burst release. After pause, and for speaker S2, who seemed to have unusually long VOTs after voiceless plosives,⁸ it was set to the onset of the second formant. The default offset of the target vowels was the offset of the first three formants. When there was a gradual and continuous transition from the vowel into the preceding or following acoustic segment, the edges of the vowel were set within the transition, their point of location determined perceptually.

First and second formant frequencies were automatically measured every 5 ms on each vowel sequence, using the ESPS function *formant*. One measure-

⁸ Long VOTs have been recently observed for at least one other native female speaker of Parisian French (see Michaud 2002:154). The influence of English cannot be discarded in either case.

ment for F1 and one for F2 were extracted at the exact midpoint of the vowels. Extreme formant frequency values were checked using both an FFT spectrum and an LPC spectrum computed over a 50 ms window centered at the midpoint of the vowel. Measurements in the two readings were then compared, and in case of important deviations, they were revised manually. Measurements from the two repetitions were then subjected to statistical analyses.

Our expectations were to find acoustic evidence for anticipatory assimilation in all word classes, and in all three speakers' readings. In other words, we expect to show a statistically significant influence of V2 on V1 with respect to F2 in all contexts, and for all speakers. The back vowel [o] is also expected to assimilate to the following final vowel in the front/back dimension. In the Southern variety, the absence of low-mid (half-open) front vowels in final open syllables (in words like *béquet*, *Bobet*), and the lack of high-mid (half-closed) vowels in closed syllables (in words such as *bosseuse*, *causeuse*, *chauffeuse*) can have an impact on the degree of vowel-to-vowel assimilation in our data. Thus each speaker's rendition and each word class have to be examined separately.

4.0 Results and discussion

Systematic variations, consistent with assimilatory phenomena referred to as VH in French, were found in the acoustic structure of V1 depending on the place of articulation of V2. However, contrary to our expectations, not every speaker showed significant anticipatory assimilation in all word sets.

Mean F2 values and corresponding standard deviations for the *été-éther*, *prêtese-prêteur*, *dénot-dévote*, and *épice-épate* sets, i.e., in which V1 was the mid front vowel [e], are shown in Table 2 and Figure 1. Each pair of bars corresponds to the set of target words indicated on the parallel axis, with the first bar representing V1, the second V2. Asterisks indicate statistically significant differences between the F2 frequencies of the two vowels, as computed in two-tailed, paired t-tests for each set⁹.

In speaker S1's readings, V1 was more [e]-like in words such as *été*, and more [ɛ]-like in *éther*. Likewise, V1 was closer to [e] than to [ɛ] in words such as *épice*, when compared to words like *épate*.

⁹ When roughly 95% of all F2 were between \pm two standard deviations from the mean, we considered them to show normal distribution. Word pairs in which measurements were impossible were discarded, which explains occasional mismatches in some t-tests' degrees of freedom across the speakers (Tables 3 and 4).

Table 3: Differences between mean F2 values, 't', and 'p' values for paired, two-tailed t-tests in four sets of target words with V1 as a high-mid vowel.

speaker/set	<i>été-éther</i>	<i>prêteur-prêteuse</i>	<i>dévoit-dévote</i>	<i>épice-éplate</i>
S1				
diff. means	42.4638	27.96	5.1667	125.93
t value	t(68) = 3.6596	t(28) = 1.174 ^{ns}	t(23) = 0.234 ^{ns}	t(47) = 6.877
p value	p = 0.0005	p = 0.2504	p = 0.8168	p = 0.0001
S2				
diff. means	60.9861	84.53	72.5	117.02
t value	t(71) = 6.003	t(29) = 5.258	t(23) = 3.507	t(47) = 10.95
p value	p=0.0001	p=0.0001	p=0.0019	p=0.0001
S3				
diff. means	22.29	-1.593	-74.875	65.89
t value	t(71) = 1.038 ^{ns}	t(31) = -0.09 ^{ns}	t(23) = -1.81 ^{ns}	t(55) = 2.259
p value	p = 0.3024	p = 0.9296	p = 0.0827	p = 0.0278

The difference of 42 Hz in the first case, and of 126 Hz in the second, was highly significant. In the *prêteuse-prêteur* and *dévoit-dévote* sets, however, the differences of 28 Hz and 5 Hz between the means were not significant. Notice, however, that these sets contained only eight and six word pairs, respectively, which means that the observed small differences could become larger when more target words examined. Speaker S2 was the only speaker whose readings reflect significant V1 to V2 assimilation in all word sets. For her, mean F2 frequencies were consistently higher before a non-low vowel than before a low vowel. Differences in F2 ranged from 61 Hz and 72 Hz, in the *été-éther* and *prêteur-prêteuse* sets, to 117 Hz in the *épice-éplate* set. A significant difference of 84 Hz was also obtained between the means in the *dévoit-dévote* set.

Similar to the other speakers, S3 pronounced V1 significantly more like [e] in words such as *épice* compared to words like *éplate*. The three other word sets, on the other hand, do not show statistically significant differences. The relatively large, -74.87 Hz difference between the means turned out to be non-significant in the *dévoit-dévote* set,¹⁰ as did the differences of -1.6 Hz and 22 Hz in the *prêteuse-prêteur* and *été-éther* sets, respectively.

The absence of V1 to V2 assimilation in the *prêteuse-prêteur* set could be due to dialectal differences, since S3, the only Southern speaker in the sample, could pronounce the final vowel identically as low-mid ([œ]) in both words of these pairs. Contrary to the Parisian pattern, V1 in her readings would then not be realized high-mid ([ø]) in closed syllables before /z/ (see 1.0.). This, however, was

¹⁰ The relatively large difference between the means, however, might call for retesting the normality of the distribution by other statistical means.

not the case. Although the above pattern is typical in her everyday vernacular, speaker S3 largely hyper-corrected her speech for the purposes of the recording: she pronounced the final vowel in words like *prêteuse* high-mid rather than low-mid, i.e., closer to the Northern rather than the Southern pattern. The lack of assimilation in the *prêteuse-prêteur* set in S1's speech (a Northern speaker) also reinforces the impression that dialectal differences alone do not explain these variations.

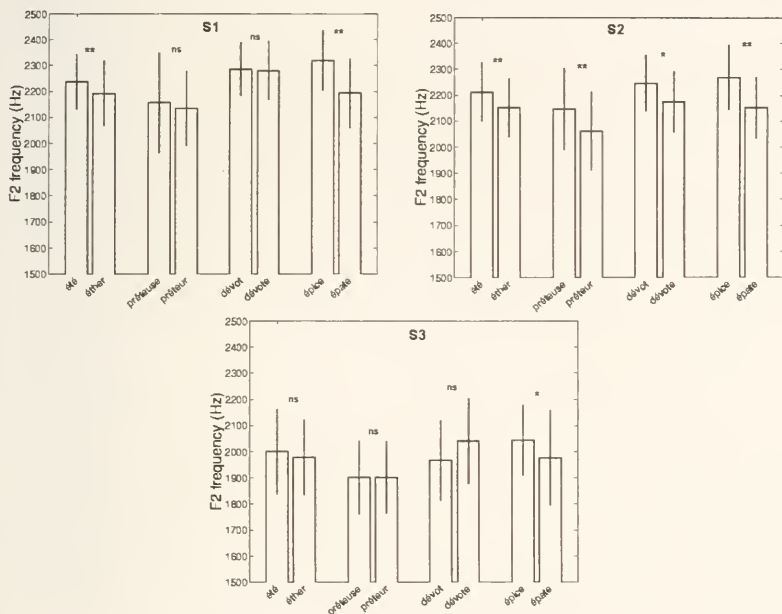


Figure 1: Mean F2 values and standard deviations for four sets of words containing a word-initial mid front vowel [e], and followed by the vowels [e]/[ɛ], [o]/[ɔ], [ø]/[œ] or [i]/[a] for two Parisian (S1 and S2) female speakers and one Southern (S3) female speaker (** significant at $p < 0.01$, * significant at $p < 0.05$, ns non-significant).

In the *dévoit-dévôte* set, our conclusions remain temporary. The fact that each speaker shows a different pattern of realization prompts us to believe that the assimilatory influence of a back rounded vowel on a high-mid front vowel is likely to involve articulatory gestures, such as rounding and height, whose acoustic correlates have yet to be investigated. Mean values of F2 for V1 are virtually identical, although statistically non-significant, in S1's speech (Δ mean = 5 Hz), they are significantly higher ($p < 0.05$) before the high-mid vowel [o] for S2, and somewhat lower before the same vowel for speaker S3. Since the place of articulation of the vowel [o] is further back than that of the vowel [ɔ], the lowering of F2 before [o] in S3's speech could be interpreted as anticipation of the backness and rounding of [o]. On the other hand, [e] becomes more [e]-like by becoming more closed and

more fronted before [o], which can be viewed as an enhancement of the vowel in its quality of a high front-vowel in S2's speech (possibly height/tense harmony).

Results for the *potée-poterne*, *poseuse-poseur*, *auto-automne*, and *notice-nota* sets containing /o/ in V1 are shown in Table 4 and Figure 2. Notice that variations in F2 exhibit the same general pattern in all speakers' renditions: F2s in V1 were systematically higher in words containing a low rather than a high mid vowel in the tonic syllable, but they were lower if V2 was an [a] as opposed to an [i]. Only two non-significant differences were obtained: one for S1 in the *poseuse-poseur* set, and one for S3 in the *auto-automne* set. The latter contained only four word pairs. Patterns of V1 to V2 assimilation were again the strongest in S2's readings. The general trend in terms of differences between the means (see Table 4) shows that V1 was more like a high-mid vowel when it was followed by a high-mid vowel, i.e., it was more [o]-like with lower F2, in words such as *potée* and *poseuse*. Also, V1 resembled more a low-mid vowel before another low-mid vowel, i.e., it was more [ɔ]-like with higher F2, in words like *poterne* and *poseur*. Thus, except for the two non-significant patterns for S1 and S3, the general assimilation pattern seems to be assimilation by backness/frontness AND height.

Table 4: Differences between mean F2 values, 't', and 'p' values for paired, two-tailed t-tests in four sets of target words with V1 as a low-mid vowel.

speaker/set	<i>potée-poterne</i>	<i>poseuse-poseur</i>	<i>auto-automne</i>	<i>notice-nota</i>
S1				
diff. means	-63.8702	-76.875	-60.8	106.75
t value	t(130) = -4.592	t(15) = -2.06 ^{ns}	t(74) = -3.979	t(63) = 3.955
p value	p = 0.0001	p = 0.0575	p = 0.0002	p = 0.0002
S2				
diff. means	-54.8433	-64.75	-53.027	39.5
t value	t(133) = -6.77	t(15) = -3.728	t(73) = -5.429	t(61) = 2.669
p value	p = 0.0001	p = 0.0002	p = 0.0001	p = 0.0097
S3				
diff. means	-29.3897	-101.625	-22.658	30.39
t value	t(135) = -2.829	t(15) = -3.645	t(75) = -1.663 ^{ns}	t(63) = 2.132
p value	p = 0.0054	p = 0.0024	p = 0.1005	p = 0.0369

In the *notice-nota* set, all speakers' F2 frequencies turned out to be higher, i.e., resembling more the low-back vowel [ɔ], when V1 was followed by [i] rather than [a]. Although this pattern is consistent with the idea of an assimilatory process, because it means that the place of articulation of V1 is probably further front when the final vowel is a high-front vowel, it goes beyond the classical definition of VH as assimilation by height. Let's suppose, for the purpose of demonstration, that VH would also be operating in the *notice-nota* set in the front/back dimension,

as it does in the three other sets. If this were true, then under the assimilatory influence of the tonic vowel [i], VH would have turned V1 into a more closed high-mid vowel in *notice*, i.e., the second formant of V1 would have been lowered, and the vowel become more [o]-like. Similarly, through VH, [a] would have made the preceding vowel a more open, [ɔ]-like low-mid vowel in *nota*, which would then have showed higher F2 frequencies. None of this is, however, the case. F2 frequencies go contrary to this hypothetical pattern: they are higher before [i] than before [a]. Thus we can conclude that the vowel is fronted, but probably not less open under the influence of [i].

Again, we might speculate at this point whether reasons for the assimilation patterns observed in the *notice-nota* set do not lie in the many factors likely to come into play in VH. Among these is the higher tongue body position, and greater lip rounding of the assimilated vowel. The latter could be due, for instance, to on-going sound change, i.e., the centralization of /o/ to /œ/ attested in the French of Île-de-France. The potentially concurrent influences of VH and the general fronting of the vowel system in middle-class Parisian French already signaled by Lennig (1978) can be difficult to sort out even in an empirical investigation (Malderez 1995, Fagyal et al. 2002).

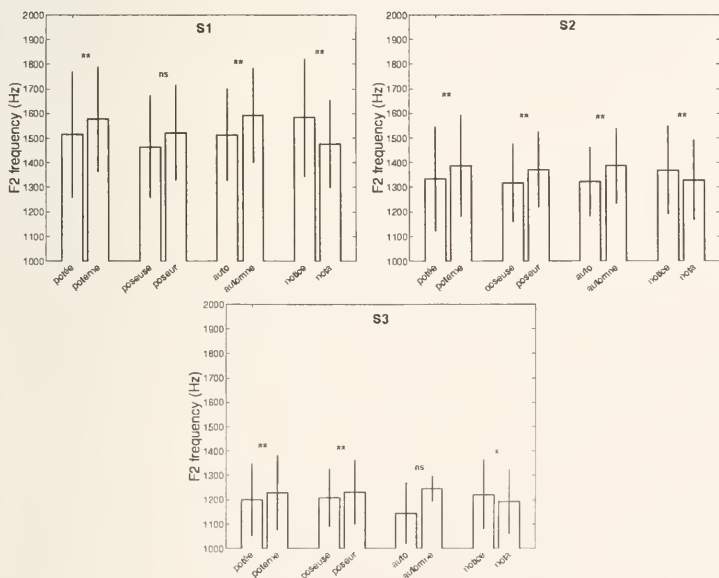


Figure 2: Mean F2 values and standard deviations for four sets of words containing a word-initial mid back vowel [o], and followed by the vowels [e]/[ɛ], [o]/[ɔ], [œ]/[œ] or [i]/[a] for two Parisian (S1 and S2) female speakers and one Southern (S3) female speaker (** significant at $p < 0.01$, * significant at $p < 0.05$, ns non-significant).

5.0 Conclusion and general discussion

In summary of this yet preliminary investigation on vowel-to-vowel assimilation referred to as 'vowel harmony' in French, one must first answer the question in the title: does VH exist in French? From an empirical point of view, the answer is: yes. Our data indicate statistically significant anticipatory assimilation in the front/back dimension in most disyllabic word sets we examined in three female speakers' speech. Therefore, according to the broadest definition of 'harmony', i.e., 'the way the articulation of one phonological unit is influenced (is 'in harmony' with) another unit in the same word or phrase' (Crystal 1997:180), the observed assimilation phenomena can be called vowel harmony.

We found systematic differences in the acoustic shape of a mid vowel contingent upon the following vowel by taking the second formant frequencies (F2) of the target vowels. For the front mid vowel, as contained in initial syllables of target words in the *été-éter*, *prêtese-prêteur*, *dévet-dévote*, *épice-épate* word sets, the results look reasonably clear: F2 is higher before a non-low vowel than before a low vowel in 58% of all cases (7 out of 12 word sets), and all speakers combined. Among the two Northern speakers, the trend is even stronger (83%). As for the back vowel in the pretonic syllables of words in the *potée-poterne*, *poseuse-poseur*, *auto-automne*, and *notice-nota* sets, one can conclude that the tendency for fronting/backing under the influence of the following vowel is even stronger: 83% of all target word sets and speakers combined, and over 90% when examined in the two Northern speakers' readings.

The second question, whether 'vowel harmony' is the right term to use with reference to the type of assimilation phenomena observed in our data, is more difficult to answer. First, even in the phonology literature, as Anderson (1980:1) pointed out more than twenty years ago, 'there is less of a consensus than meets the eye as to just what the characteristics are that set vowel harmony apart from other types of rule'. According to Hyman's (2002) more recent reformulations — 'What is vowel harmony? Where does it come from? How is it the same as or different from other things?' — the issue is far from being resolved. In lack of general consensus on how VH, metaphony, umlaut and possibly other vowel assimilation phenomena should be separated from each other in a precise way, we retain the following arguments for the purposes of our investigation.

According to the narrowest definitions of vowel assimilation phenomena inherited from nineteenth century historical phonetic studies, we would reserve the term 'vowel harmony' for phonological processes well-known from languages such as Turkish and Hungarian. That is, VH is a process responsible for the selection of an allomorph for a given base word from two or three candidates. Whether the directionality of this process is an artifact of sampling in phonological studies, or a universal bias in processing or perception, as Hyman (2002) seems to argue, still awaits an answer. Recent empirical studies indicate that front/back assimilation can represent a production advantage at the word level (Cole et al. 2002), and that anticipatory assimilation does not conflict with carry-over vowel harmony in languages such as Turkish (Inkelas et al. 2001). Until these factors are sorted out,

however, directionality can be used to distinguish between VH, in the narrowest sense of the term, and metaphony. Vowel assimilation observed in French seems to be better described in terms of the latter. Metaphony phenomena triggered by final high vowels are well-known from the history of French, and are abundant in other Romance languages as well (e.g., Posner 1997, Hualde 1989). Thus in light of the Romanist tradition that reserved the term 'metaphony' for anticipatory assimilatory changes triggered by non-low vowels at the word level, we suggest that vowel assimilation observed in our data are better understood as a particular — although certainly not classical — case of metaphony, rather than VH.

As for the final question, whether anticipatory vowel assimilation in French operates in specific morphological environments or also in words not tied to derivational or inflectional paradigms, we can only speculate about the answer. Based on the non-significant influence of V2 on V1 in the *prêteuse-prêteur* set (for speakers S1 and S3) and the *poseuse-poseur* set (for speaker S3), containing only semantically related word pairs, we can speculate that semantic relatedness might not be a determining factor. This impression is further supported by word sets that contain a majority of morphologically and semantically unrelated word pairs (*épice-épate*, *notice-nota*), and yet, indicate strong anticipatory assimilation between the two neighboring vowels. Future studies might investigate this in greater detail by separating, and individually testing each of these word pairs.

APPENDIX: Total number of word pairs (N=136) subdivided in sets according to the quality of their first (V1) and second (V2) vowels. The 17 pairs containing an initial closed syllable, shown in *italic*, were not included in the analysis. Words in **bold** are pronounced differently in the two dialects.

Label of the set	V1	V2	Word pairs in the set
été — éther	e	e/ɛ	ailé-ailette, aîné-aïnesse, <i>berger-bergère</i> , béquet-Beckett, clairet-clairette, défait-défaite, épée-épaisse, <i>fermé-fermette</i> , fléché-fléchette, <i>germé-Germaine</i> , méfait-méfaire, mémé-mémère, pépé-pépère, rainé-rainette, réglé-réglette, rêvé-rêveil, sellé-sellette, séné-séneestre, télé-télex, terrer-terrestre, <i>verger-Vergès</i> ;
prêteuse — prêteur	e	ø/œ	<i>chercheuse-chercheur</i> , fraiseuse-fraiseur, gêneuse-gêneur, payeuse-payeur, plaideuse-plaideur, prêcheuse-prêcheur, rêveuse-rêveur, veilleuse-veilleur;
dévoit — dévote*	e	o/ɔ	écot-école, épaule -époque, Esso-essor, étai-étouffe, <i>recto-rectum</i> , vélo-vélocé;
potée — poterne†	o	e/ɛ	Bobet -bobèche, <i>bordée-bordel</i> , bosser-bossette, broché-brochette, causer-causette, chaussée-chaussette, chorée-chorège, clocher-clochette, cocher-cochère, coder-codex, coller-collecte, coquet -coquette, <i>cordée-cordette</i> , <i>cornée-corneille</i> , <i>corsé-corsaire</i> , <i>corvée-corvette</i> , croquer-croquette, forer-foraine, <i>forcer-forceps</i> , <i>former-formel</i> , fossé-fossette, <i>gauffré-gauffrette</i> , <i>lorgner-lorgnette</i> , mauvais -mauvaise, mollet -mollesse, moquer-moquette, motet -motel, noter-notaire, orée-oreille, oser-oseille, pocher-pochette, pommer-pommette, protégée-prothèse, roquet -roquette, rosé-rosette, sauter-sauterne, sommet -sommeil, sonner-sonnette, tollé-Tolède, tonner-tonnelle, violet -violette;
poseuse — poseur‡	o	ø/œ	bosseuse -bosseur, causeuse -causeur, chauffeuse -chauffeur, chômeuse -chômeur, colleuse -colleur, croqueuse -croqueur, donneuse -donneur, fauteuse -fauteur, faucheuse -faucheur, foreuse -foreur, fraudeuse -fraudeur, moqueuse -moqueur,

Label of the set	V1	V2	Word pairs in the set
			noceuse-noceur, <i>porteuse-porteur</i>, rôdeuse-rôdeur, sauteuse-sauteur, sonneuse-sonneur, trotteuse-trotteur, veilleuse-veilleur, voleuse-voleur;
auto — automne	o	o/ɔ	coco-cocotte, mono-monocle, sono-sonore;
épice — épate	e	i/a	bêtise-bêtasse, béquille-bécane, dédit- dédale, éthyle-étal, ferrite-ferraille, fléchir- fléchage, hélice-hélas, Messire-message, réglisse-réglage, sénile-sénat, série-sérail, terrine-terrasse, vicilli-vieillard;
notice — nota	o	i/a	bobine-bobard, choriste-chorale, coquine- cognac, colline-collage, Corinne-coral, Dorine-dorade, <i>forcir-forçage</i> , mollir- mollah, Monique-monarque, motif-motard, otite-otage, postiche-postal, potiche- potache, protide-protase, Rosine-rosace, sonie-sona;

*Due to the strict observance of the closed-syllable adjustment rule, in Southern French dialects, V2 in *épaule* is low-mid, [ɔ], underlyingly.

†All final front mid vowels are expected to be high-mid, [e], in the South.

‡In Southern French dialects, regardless of the type of the coda consonant, all non-front mid-vowels are low-mid, [œ] or [ɔ], in closed syllables.

REFERENCES

- ANDERSON, Stephen R. 1980. Problems and perspectives in the description of vowel harmony. In: Vago (ed.), 1-48.
- BATTYE, Adrian, Marie-Anne HINTZE, & Paul ROWLETT. 2000. *The French Language Today*. London and New York: Routledge.
- BLAMPIN, Daniel (ed.). 1997. *Le français en Belgique: une langue, une communauté*. Louvain-la-Neuve: Duculot.
- BUSSMANN, Hadumod. 1996. *Routledge Dictionary of Language and Linguistics*. London, New York: Routledge.
- CASAGRANDE, Jean. 1984. *The Sound System of French*. Washington D. C.: Georgetown University Press.
- COLE, Jennifer, Alina KHASANOVA, & Gary DELL. 2002. Evidence of a production basis for front/back vowel harmony. Paper presented the LabPhon 8 Meeting. Haskins Laboratories, CT.
- CRYSTAL, David. 1997. *A Dictionary of Linguistics and Phonetics*. 4th edition. Oxford, Cambridge: Blackwell Publishing.
- DELL, François. 1972. *Les règles et les sons: Introduction à la phonologie générative*. Paris: Hermann.
- FAGYAL, Zsuzsanna, Samira HASSA, & Fallou NGOM. 2002. L'opposition [e]-[ɛ] en syllabes ouvertes de fin de mot en français parisien: étude acoustique préliminaire. *Actes des Journées d'Etudes sur la Parole, Nancy, France*, 165-8.
- FOUCHÉ, Pierre. 1952. *Phonétique Historique du Français*. Paris: Klincksieck.
- . 1956. *Traité de prononciation française*. Paris: Klincksieck.
- FOUGERON, Cécile, & Caroline SMITH. 1999. French. In: *Handbook of the International Phonetic Association* (Cambridge University Press, Cambridge, UK), 78-81.
- GADET, Françoise. 1997. *Le français ordinaire*. Paris: Armand Colin.
- GRAMMONT, Maurice. 1914[1939]. *Traité de phonétique*. Paris: Delagrave.
- HOCK, Hans Henrich. 1991. *Principles of Historical Linguistics*. Berlin: Mouton de Gruyter.
- HUALDE, José Ignacio. 1989. Autosegmental and metrical spreading in the vowel-harmony systems of Northwestern Spain. *Linguistics* 27:5. 773-805.
- HYMAN, Larry, M. 2002. Is there a right-to-left bias in vowel harmony? Paper presented at the 9th International Phonology Meeting on 'Structure and Melody'. Vienna (Austria).
- INKELAS, Sharon, Alan C.L. YU, Jonathan BARNES, Jeffrey GOOD, Darya KAVITSKAYA, Orhan ORGUN, & Ronald SPROUSE. 2001. Stress and vowel-to-vowel coarticulation in Turkish. Paper presented at the Annual Meeting of the Linguistics Society of America. Washington D.C.
- LANDICK, Marie. 1993. L'Harmonie vocalique: Prescription or description. *The Linguist* 32:1. 34-5.
- . 1995. The mid-vowels in figures: Hard facts. *The French Review* 69:1, October. 88-102.

- LENNIG, Matthew. 1978. Acoustic measurement of linguistic change: the modern Paris vowel system. Department of Linguistics, University of Pennsylvania, MS.
- LÉON, Pierre. 1966. *Prononciation du français standard. Aide-mémoire d'Orthoépie*. Paris: Didier.
- MALDEREZ, Isabelle. 1995. Contribution à la synchronie dynamique du français contemporain: le cas des voyelles orales arrondies. Department of Linguistics, Université de Paris 7, MS.
- MALMBERG, Bertil. 1966. *La Phonétique*. Paris: Presses Universitaires de France.
- . 1969. *Phonétique française*. Malmö: Hermonds.
- . 1974. *Manuel de phonétique générale: Introduction à l'analyse scientifique de l'expression du langage*. Paris: Editions A. & J. Picard.
- MARTINET, André. 1945. *La prononciation du français contemporain*. Genève: Droz.
- MCCORMICK, Mary S. 1982. Vowel Harmony and umlaut: Implications for a typology of accent. Department of Linguistics, Cornell University, MS.
- MICHAUD, Alexis. 2002. Pourrais-tu enregistrer un corpus pour moi? Pour une charte de qualité des corpus. *Actes des Journées d'Etudes sur la Parole, Nancy, France*, 153-6.
- MORIN, Y. C. 1971. *Computer Experiments in Generative Phonology. Low-Level French Phonology*. Natural Language Studies 11. Department of Computer and Communication Sciences, The University of Michigan, Ann Arbor.
- OHALA, John J. 1994. Towards a universal, phonetically-based theory of vowel harmony. *Proceedings of the ICSLP, Yokohama*. 10-14.11 - 10-14.14.
- POSNER, Rebecca. 1997. *Linguistic Change in French*. Oxford: Oxford University Press.
- SELKIRK, Elizabeth. 1972. *The Phrase Phonology of English and French*. Department of Linguistics, MIT, MS.
- TRANEL, Bernard. 1987. *The Sounds of French: An Introduction*. Cambridge: Cambridge University Press.
- TRASK, R. L. 1996. *A Dictionary of Phonetics and Phonology*. London, New York: Routledge.
- VAGO, Robert M. (ed.). 1980. *Issues in vowel harmony*. (CUNY Linguistics Conference on Vowel Harmony 1977, City University of New York. Studies in Language companion series, 6.) Amsterdam: John Benjamins.
- WALKER, Douglas. C. 2001. *French Sound Structure*. Calgary: University of Calgary Press.

ON THE NATIVIZATION OF ENGLISH IN CHINA*

Liwei Gao

University of Illinois at Urbana-Champaign
liweigao@students.uiuc.edu

This paper presents a data-based analysis of the nativization of English used in the Chinese context, particularly on the discursive level. The paper argues that although in China English functions primarily as a performance or exonormative variety, it has been nativized within the Chinese culture and society. The nativization lends support to the claim that a Chinese variety of English is emerging, and manifests Chinese-English bilinguals' linguistic creativity. This paper also briefly discusses the implications for the selection of appropriate models of English to teach in China.

I. Introduction

It is now well recognized that English has reached virtually every corner of the globe (e.g., Crystal 1997; McArthur 1987, 1993). There is also a general consensus that there are more speakers of English in the world who use English as an additional language in a bi- or multilingual context than in a monolingual context (e.g., Hung 2002, Kirkpatrick & Xu 2002). It is therefore no exaggeration to claim that English has obtained the status of an international language, world language, or *lingua franca* (e.g., Brutt-Griffler 2002, B. Kachru 1996, Pakir 2001). In China, it is generally claimed that there are now approximately 200 to 300 million English speakers, even though they use English with various degrees of proficiency (e.g., Dzau 1990, McArthur 1992, Zhao & Campbell 1995). As such, this is the largest number among countries that belong to the Expanding Circle¹ (Crystal 1985). Furthermore, in China, English is used in both INTER- and INTRANational domains.²

* An earlier version of this paper was written for the seminar on Sociolinguistics of World Englishes offered at the University of Illinois at Urbana-Champaign in the fall semester of the year 2000. I would like to express my deep gratitude to Professor Braj B. Kachru for his valuable comments, critiques, and suggestions from which this paper has immensely benefited. Portions of this paper were read at the 2002 annual meeting of the Central States Anthropological Society held at Michigan State University, 7-10 March. Thanks also go to those who kindly offered their comments and suggestions at the meeting, particularly to James Stanlaw and Nobuko Adachi. Yet, any shortcomings in this paper are solely my own responsibility.

¹ English-speaking countries are often categorized as three concentric circles in the Kachruvian model of World Englishes (1985): the Inner Circle, the Outer Circle, and the Expanding Circle. The Inner Circle contains countries where English is used as a native language, such as Australia, New Zealand, and the US. The Outer Circle countries are those in which English has been nativized and institutionalized for both INTER- and INTRANational communication, such as India, Malaysia, and Singapore. In the Expanding Circle are countries where English is learned and used primarily as a foreign language, such as China, Japan, and the Republic of Korea.

² Gao (2001) provides an updated survey of the functions of English in China. He notes that English in China functions in INTRA- as well as INTERNATIONAL domains. The international use is primarily found in

With the establishment of the first British trading post in Guangzhou (Canton) in the late seventeenth century (e.g., Bolton 2000, 2002; Pride & Liu 1988), English made its first entry into China. Then with the opening-up policy initiated by the Chinese government in the late 1970's, English has surpassed all other foreign languages, for example, Russian and Japanese, in fighting for the linguistic market and consequently has become the most popular foreign language in China. It has turned into the language that promises, among other things, job opportunity, upward 'social and economic mobility' (Zhao & Campbell 1995:385), and social prestige. According to Bolton (2002), with China's admission into the World Trade Organization (WTO) in November 2001 and the selection of Beijing, China's capital city, as the host city for the 2008 Olympic Games, the craze for English seems to have reached a new peak in China with, for example, the general public, educators, and government policy-makers. The overwhelming enthusiasm for English learning is reflected even in a documentary film, *Crazy English*, produced and directed by Yuan Zhang, which reports on how an English teacher named Yang Li gained a pop star type of status by teaching English nationwide.

In one of the earlier studies on the nativization of English in China, Cheng (1992) labels English used in the Chinese context as 'Sinicized English'. Others (e.g., Zhang 1997) term the same variety of English 'China English'. It is not crucial which terminology is used. What is significant is that these terms are used with reference to the nativization of English in lexis, syntax, phonology, discourse, etc. This, in turn, has substantial bearing on the issue of the nativization of English in different sociocultural contexts (e.g., B. Kachru 1965, 1983). While there has been much documentation on the lexical acculturation of English in the Chinese context (e.g., Cheng 1992, Gao 2001), research is relatively rare on discursive, syntactic, or phonological nativization.³ This paper examines the nativization of English in China from a primarily discursive perspective, where the term 'discourse' is operationally defined as the linguistic unit that is larger than a lexical item or a phrase.

In China, English is generally considered to be a norm-dependent or ex-normative variety. In other words, it mostly functions as a foreign language.⁴ This assumption is essentially a consequence of the limited penetration of English in Chinese society, in terms of both the range and depth, which is closely related to China's social and historical background. It is true that in the Chinese context, the degree to which English is nativized is not as large as in the Outer Circle countries, where English has now been institutionalized and also serves as a do-

foreign trade, international business, tourism, science, and technology. Intranational domains mainly cover media, education, and translation. Gao also points out that the INTER- versus INTRANational dichotomy is not meant to be clear-cut.

³ Cheng (1992) focuses on the examination of lexical acculturation in the political arena. Gao (2001) surveys lexical acculturation in a variety of fields, for example, in trade and economy, education, and traditional customs and arts.

⁴ This is largely true. However, as Pride & Liu (1988) note, based on the amount of English used in the Chinese media, the degree to which Chinese policies favor English, and the favorable attitude toward English among the general public, English may be rightly regarded as a second language in China, ranking only second to Chinese.

mestic *lingua franca*. It is basically due to these realities that some scholars, both in China and abroad, are still hesitant to acknowledge that English in the Chinese context has been nativized. This is, however, an attitudinal issue. In light of the theoretical framework of B. Kachru (1980, 1992a), it is only natural that the nativization of English should have taken place in the Chinese context.⁵

2. Theoretical framework

This study is conducted essentially within the framework established by B. Kachru (1980, 1992a, 1986a, 1986b, 1987a, 1987b) and Y. Kachru (1987, 1995). According to this general framework, whenever English is used in a country or region where it is not the native language, be it for science, technology, literature, or modernization, it undergoes a process of 'reincarnation' (Kachru 1992a), both linguistic and cultural. B. Kachru (1987) considers most such changes functionally determined, given the fact that a nonlocalized variety of English is not capable of adequately expressing what is unique to a culture where English is not the native language, whereas the nativized variety can. As B. Kachru observes (1992b), a linguistic convention of language use necessarily reflects the conventions of culture.⁶

In addition, B. Kachru (1987a) notes that bilinguals' creativity also plays a significant role in discursive nativization of English in its nonnative linguistic ecology, which may be represented by distinctive text designs, stylistic conventions, culture-specific speech acts, and traditional thematic ranges of the discourse.⁷ B. Kachru (1987b) further points out that discursive nativization results mostly from bilinguals' strategic transfer of discursive features from their dominant language to the less used language in the linguistic realizations of the underlying thought patterns. The nativized English discourse generally does not conform to the recognized canons established in the native variety of English. Nevertheless, such nativization, according to B. Kachru (1987a), is an inevitable consequence of the unprecedented spread of English throughout the world.

Similarly, as Y. Kachru (1987) observes, there exist distinct discourse patterns and discourse strategies in different languages and cultures. Moreover, although there is no difference in underlying cognitive processes among various peoples, there do exist differences in conventionalization of discursive patterns. Elsewhere, Y. Kachru (1995) observes that nonnative varieties of English usually feature some writing patterns and rhetorical structures that differ from those in the

⁵ The nativization of English investigated in this paper is a consequence of contact between Chinese and English in relatively recent historical periods. Earlier stages of contact between Chinese and English have produced, among other things, English borrower words from Chinese, e.g., *litchi*, *chow mein*, *kowtow*, *kung fu*, *typhoon*, and *mahjongg*. According to McArthur (1992), English has nearly 1,000 such Chinese loan words.

⁶ Zuo (2001) holds a similar viewpoint that cultural differences are inevitably exhibited in linguistic differences given that language is an integral component of culture.

⁷ Zhang (2002) examines the linguistic creativity of bilinguals expert in both Chinese and English by analyzing the English novella *In the Pond*, which was written by the Chinese writer Jin Ha. In this article Zhang discusses the bilinguals' creativity on both lexical and discursive levels. Kirkpatrick & Xu (2002) also contains a brief discussion of discursive nativizations of English used in the Chinese context.

English writing by people in the Inner Circle countries. Related to this issue, Y. Kachru (2001) further notes that in order to gain discourse competence in World Englishes, viz., in order to be successful in conducting crosscultural communication that involves different varieties of English, speakers of English must strive to be familiar with, and also be sensitive to, the distinctive discursive features in different varieties of World Englishes.

3. Data collection

In this study, most of the samples examined were collected from *Beijing Review*, one of the major English magazines in China. *Beijing Review* was first published in 1958 and is now published weekly. A small number of sample sentences were also collected from *China Daily*, currently one of China's most widely read English newspapers. *China Daily* was started in 1981. It is generally assumed that the English used in these two publications is representative of the Chinese variety of English. Although native speakers of English were, and still are, on the editorial staff of both publications, a number of English usages still carry with them Chinese features.

The data analyzed in this study were collected from the articles published in *Beijing Review* in 1968 and also those published in *China Daily* in 2000. The motivation for the selection of data from two different historical periods is to show that discursive nativization of English takes place in different sociopolitical contexts, even though different eras may be characterized with dissimilar types of nativization. In addition, the inclusion of data from both journal and newspaper articles is to demonstrate that the nativization of English in China is not found simply in one type of publication. Other than these considerations, the collection of data was random.

4. Discursive nativization of English in China

4.1 General overview

It is not clear whether the typical discourse organization pattern of the native variety of English has been localized in the Chinese context. For one thing, according to Myers (2000), there exists conflicting evidence regarding the differences and similarities between the discourse organization patterns of English and Chinese. A considerable number of researchers (e.g., Jia & Cheng 2002; Kaplan 1966; Kirkpatrick 1991, 1995; Young 1982; Zuo 2001) suggest that Chinese writers do not follow the linear thought pattern of beginning a discourse with a topic sentence, followed by supporting sentences that precede the conclusion, which is characteristic of Western cultures. On the contrary, the discourse organization pattern in the Chinese language is considered to be circular or indirect. However, this assumption is challenged by other scholars. For example, Mohan & Lo (1985) contend that the Chinese writing taught in schools fosters a direct writing style. They have presented examples of direct writing from classical Chinese literature to support their argument.

Unlike the nativization of discourse organization patterns in Chinese English, which is still a somewhat controversial issue, the nativization of discourse strategies is less ambiguous (e.g., Kirkpatrick & Xu 2002, Myers 2000). It is represented by, but not limited to, the extended use of localized speech acts, rhetorical devices, idiomatic expressions, or political slogans. These nativized discourse strategies, which mostly result from the transfer of Chinese discourse strategies to Chinese English discourse, are all loaded with Chinese characteristics. These nativized discursive strategies function, to some extent, as markers of China English discourse.

4.2 Data analysis

In this section, the paper identifies and examines different types of discursive nativizations that have taken place in Chinese English, which convey peculiar sociocultural information about the Chinese society. These nativized usages are not immediately transparent to those native speakers of English⁸ who are not familiar with Chinese culture. As B. Kachru (1990:165) points out, discursive nativization of English is a phenomenon where a text is overloaded with cultural presuppositions, which, as a result, 'demands a serious cultural interpretation'.

It is worth noting that the localized conventions to be analyzed in this study are usually used more in written discourse than in the spoken variety of China English. Moreover, since the bulk of the sample sentences and paragraphs examined were taken from articles published in 1968, during which China was in the midst of the Great Proletarian Cultural Revolution, a ten-year long political movement (1966-1976), certain types of discursive nativizations discussed in this article may not be found in the English writing from other historical periods.

For convenience in reading, in this paper, the texts that feature discursive nativizations are given in italics, whereas most other typographical features of the original text, such as the sometimes unusual use of capitalization and punctuation marks, are preserved.

4.2.1 Re-creation of local speech acts

4.2.1.1 Wishes

Openly expressed wishes for good health or longevity, both of which are much valued in traditional Chinese culture, are seldom found in the writing of native speakers of English. Such speech acts are not part of their ordinary linguistic behavior, although certain other types of wishes are, for example, the wish for good luck or success. In contrast, in the Chinese society, especially during the Great Proletarian Cultural Revolution, extended wishes for good health or longevity, especially those to Mao Tse-Tung, were frequently made by virtually all Chinese people, old and young alike. In other words, there is a discrepancy here in terms of the thematic range found in the speech act of wishing. Examples of wishes for good health or longevity in the Chinese English writing are provided in (1-3).

⁸ In recent years the construct 'native speaker' has been subject to scrutiny and has become the object of heated debates. For a thorough discussion of this issue, see, for example, Paikeday (1985). In this paper 'native speaker' is used merely as a working concept.

- (1) "Wishing Your Excellency life-long good health."
(From *Beijing Review*, January 3, 1968)
- (2) *Revolutionary People of the World Wish a Long, Long Life to Chairman Mao!*
(From *Beijing Review*, January 12, 1968)
- (3) "Long live Mao Tse-tung!" This is confidence in victory.
(From *Beijing Review*, January 12, 1968)

In certain cases, the wishes are overly lengthy, which, to some extent, reflects the Chinese people's fervent love and respect for Mao Tse-Tung on the one hand, and the fervidness of the political movement itself on the other. Such nativized speech acts usually strike native speakers of English as odd because many of them are not familiar with either the fervidness of the political movement during the Great Proletarian Cultural Revolution or the different thematic ranges of the speech act of wishing in the Chinese language. Examples of excessively lengthy wishes are given in (4).

- (4) They wave their red-covered copies of *Quotations From Chairman Mao Tse-tung* and shout, "Long live Chairman Mao!" "A long, long life to Chairman Mao!" "Long live the victory of Chairman Mao's proletarian revolutionary line!" "Long live the victory of the great proletarian cultural revolution!" "Long live the invincible thought of Mao Tse-tung!" and "We wish Chairman Mao a long, long life!"
(From *Beijing Review*, January 3, 1968)

On the lexical level, in (4), the expression 'red-covered' is a loan translation⁹, which has undergone semantic nativization¹⁰ by means of a semantic shift, more exactly, semantic extension, given that in the Chinese context this lexical item has acquired a sense not available in the native variety of English. Generally speaking, a thorough and accurate understanding of such loan translations presupposes knowledge of Chinese culture. The term 'proletarian' has also been semantically nativized, this time by means of semantic amelioration. In native varieties of English, the word 'proletariat' denotes 'the class of industrial wage earners who must earn their living by selling their labor' (Anderson, Fortson, Kleinedler, & Schonthal 2001:672), in other words, people with a low social and economic status. However, in (4), the word 'proletariat' does not refer to people from the low social class, although their economic status may not be as high. On the con-

⁹ The majority of the acculturated English lexicon in China English is loan translations, word-by-word translations of Chinese expressions. These nativized English expressions are used mostly because no English expressions in the native variety adequately or precisely match both the denotation and connotation of the Chinese expressions. Some of these loan translations have already developed their own abbreviations in the Chinese context, which bespeaks, among other things, Chinese-English bilinguals' linguistic creativity.

¹⁰ Kachru (1992a) observes that when lexical items are nativized, their denotation or connotation is usually either extended or restricted. Cheng (1992) also notes that in different societies words may have different connotations. On the same issue Zhou & Feng (1987) identify four different situations, i.e., semantic extension, reduction, amelioration, and deterioration. This proposal provides a helpful approach to the analysis of semantic shift. It is obvious that these four situations of semantic shift are not mutually exclusive.

trary, the 'proletariat' refers to the leadership in the Chinese revolution and social life.

In addition to the wish for good health or longevity, wishes connected with certain other aspects of the Great Proletarian Cultural Revolution are also frequently found in the articles published in the 1968 issues of *Beijing Review*. In (5) below is an example of a wish for the Revolution itself, given that in the Chinese context, particularly during the Great Proletarian Cultural Revolution, the red flag symbolizes, among other things, revolution.

- (5) *May the red flag of the ChingKang Mountains, raised by Chairman Mao, fly for ever!*

(From *Beijing Review*, January 12, 1968)

In (6) is an example of a wish that is very often considered to be made by Mao Tse-Tung himself. This time it is a wish for more glorious revolutionary achievements. These wishes usually strike native speakers of English as strange too.

- (6) *"Carry the Revolutionary Tradition Forward, May You Gain Still Greater Glory."*

(From *Beijing Review*, January 12, 1968)

On the lexical level, the expression 'red flag' is also a loan translation. This term has undergone semantic nativization by means of a semantic shift as well. In the native variety of English, the expression 'red flag' usually simply connotes the notion of 'stop'. In China English, the connotation of 'red flag' includes, among other things, revolution and the rule by the Chinese Communist government, given that, for one thing, the 'red flag' is China's national flag. In this sense, in (5), the term 'red flag' is not only semantically extended but also semantically ameliorated.

4.2.1.2 Indirect curses

In addition to using terse expressions to curse, as is usually the case in cultures where English is the native language, Chinese people also commonly use complete or partial sentences for this speech act. This nativized speech act is also observed in some of the articles in the 1968 issues of *Beijing Review*. Furthermore, in these articles the curses are usually expressed in a roundabout manner. Therefore, I call them 'indirect curses'. Moreover, these curses are mostly expressed in the form of quasi-proclamatory statements. Examples are provided in (7-10), which, in a sense, curse the United States, Great Britain, and the then Soviet Union respectively.

- (7) *"The US imperialists and all other such vermin have already created their own grave-diggers; the day of their burial is not far off."*

(From *Beijing Review*, January 12, 1968)

- (8) This "package plan," described by the British ruling circles as a heart-breaking decision, marks the fact that *British imperialism indeed is on its last legs.*

(From *Beijing Review*, January 26, 1968)

- (9) Soviet Revisionists' Plot to Call Counter-Revolutionary International Meeting Can Only *Speed Their Own Doom*.

(From *Beijing Review*, January 12, 1968)

4.2.2 Employing culturally loaded rhetorical devices

4.2.2.1 Analogy

The Chinese have been fondly employing analogy as a means of argumentation throughout their history of civilization, which is, to some degree, distinctive of Chinese culture (e.g., Myers 2000). When Chinese bilinguals write in English, they naturally make use of this rhetorical device as well, as they do when they write in Chinese. The use of analogy in Chinese English is exemplified in (10-11). In (10), the importance of Mao Tse-Tung's thought to the revolution is compared to both that of the helmsman to sailing and the sun to living things. In (11), the importance of the Communist Party to the revolution is compared to that of water to fish and also vines to melons.

- (10) *Sailing the seas depends on the helmsman, all living things depend on the sun for their growth... making revolution depends on Mao Tse-Tung's thought.*

(From *Beijing Review*, January 3, 1968)

- (11) *Fish can't live without water, melons can't thrive off their vine, the revolutionary masses cannot do without the Communist Party, Mao Tse-tung's thought is the never-setting sun.*

(From *Beijing Review*, January 3, 1968)

4.2.2.2 Symbolism

Chinese writers also frequently make use of a plethora of culture-specific symbols in their Chinese writing, which is yet another feature that, to a certain degree, marks Chinese discourse. It is not unusual for Chinese bilinguals to transfer this rhetorical device to their English writing. Examples that feature such a transfer are provided in (12-14). For instance, in (12), the use of the phrase 'the red sun' is characteristic of Chinese culture, where the sun is considered to be one of the greatest things in the world. Its symbolism includes, but is not limited to, power, righteousness, long life, and source of strength.

- (12) Chairman Mao Tse-tung, ... *the red sun* that shines most brightly in our hearts, receiving revolutionary fighters on December 31, 1967 in Peking's Great Hall of the People.

(From *Beijing Review*, January 3, 1968)

In (13), the verses in Mao Tse-Tung's poem symbolize the imminent burst of revolution throughout the world.

- (13) "*The Four Seas are rising, clouds and waters raging, the Five Continents are rocking, wind and thunder roaring,*" wrote Chairman Mao in one of his poems.

(From *Beijing Review*, January 12, 1968)

On the lexical level, in (14), the term 'red book' has also undergone semantic nativization by means of a semantic shift.

- (14) The red book of Mao Tse-tung *sheds its rays* everywhere. The proletarians read it and draw courage to fight the oppressors.

(From *Beijing Review*, January 3, 1968)

The expression 'red book' here does not simply refer to a book that is red. Instead, it specifically refers to the much-revered little red book that collects Mao Tse-Tung's instructions and advice. In this sense, this term has undergone both semantic reduction and semantic amelioration.

4.2.2.3 Metaphor

Myers (2000) notes that metaphor is also commonly used in the writings of Chinese people. In addition, Tang (1997) observes that Chinese philosophical writings made more frequent use of metaphors to make points than Western philosophical works did. Then, it may be the case that Chinese bilinguals have utilized this rhetorical device in their English writing. The examination of the articles in *Beijing Review* confirms this hypothesis. Examples of the use of metaphors are provided in (15-18). In (15), Mao Tse-Tung's thought is compared to 'the great banner'; in (16), Mao Tse-tung is compared to 'the beacon' for the people in the whole world; in (17), the raging of war is compared to 'a prairie fire'; and in (18), reactionaries are compared to 'paper tigers'.

- (15) Let us cheer: The world has entered the new era in which *Mao Tse-tung's thought is the great banner*.

(From *Beijing Review*, January 3, 1968)

- (16) Chairman Mao, *You Are the Beacon* for the People of the World.

(From *Beijing Review*, January 3, 1968)

- (17) *Flames of People's War Raging Like a Prairie Fire* in Thailand.

(From *Beijing Review*, January 12, 1968)

- (18) *All reactionaries are paper tigers*. In appearance, the reactionaries are terrifying, but in reality they are not so powerful.

(From *Beijing Review*, January 19, 1968)

4.2.2.4 Parallelism

Parallelism is also a rhetorical device used regularly in the writings of Chinese people. It is mostly utilized to highlight a certain notion or concept. According to Di (2001), the extensive use of parallelism by Chinese people is, to a great extent, attributed to the influence from the *yin-yang* thinking in Daoism, one of the most influential religions practiced in China.¹¹ In their English writing, Chinese bilinguals usually avail themselves of this rhetorical device as well. In contrast, parallel expressions are relatively rare in the average prose written by native speakers of English. Examples of the use of parallelism in Chinese English are provided in

¹¹ Di (2001) analyzes the influence from the *yin-yang* thinking upon English writing by native speakers of Chinese. In so doing she identifies the impact, which is frequently represented by noncolloquial usages, not only on the discursive or rhetorical level, but also on the lexical and/or syntactic level.

(19-22). In (19), (21), and (22), the parallel linguistic units are sentences, whereas in (20), the parallelism is represented by noun phrases.

- (19) *On buses and trolley buses people animatedly debate how cadres are to be "liberated," or what are the criterions of a good Party member of Chairman Mao...*

On the threshing floors of people's communes bespectacled grandfathers are reading Serve the People word by word with their grandchildren.

On the sea and in the air the People's liberation Army uses quotations from Chairman Mao to direct its patrols, training and preparations against war.

(From *Beijing Review*, January 19, 1968)

- (20) *Our great teacher, great leader, great supreme commander and great helmsman* (Used to refer to Mao Tse-tung)

(From *Beijing Review*, January 19, 1968)

- (21) *"Chairman Mao, you are the great savior of mankind!" "Chairman Mao, you are the greatest Marxist-Leninist of our time!" "Chairman Mao, you are the beacon for the people of the world!"*

(From *Beijing Review*, January 3, 1968)

- (22) *"Long live Mao Tse-tung!" This is a military pledge. "Long live Mao Tse-tung!" This is the source of strength. "Long live Mao Tse-tung!" This is confidence in victory. "Long live Mao Tse-tung!" This is the cheer of happiness. "Long live Mao Tse-tung!" These are words full of respect and love.*

(From *Beijing Review*, January 12, 1968)

4.2.3 Employing language-specific idioms and slogans

The Chinese language is rich in idiomatic expressions and slogans. These idioms and slogans are full of cultural connotations that are not easily decipherable for those native speakers of English who are not familiar with Chinese culture. When Chinese people write in English, they, not surprisingly, will also use these translated sayings and idioms. These formulaic expressions reflect, among other things, ways of life, beliefs, and wisdoms of the Chinese people, all of which are in one way or another different from those of the people in the Inner Circle countries. Examples of this type are given in (23-26).

- (23) *A hundred flowers bloom and a hundred schools of thought contend.*

(From *China Daily*, November 24, 2000)

- (24) *The bells for the new century are soon to peal.*

(From *China Daily*, November 24, 2000)

- (25) *Growing is a painful process, but one that is necessary for survival in the future.*

(From *China Daily*, November 20, 2000)

- (26) Every Communist must grasp the truth, "*Political power grows out of the barrel of a gun.*"

(From *Beijing Review*, January 3, 1968)

The discussion of the above discursive nativizations found in Chinese English is by no means intended to indicate that these discourse strategies are used EXCLUSIVELY by Chinese-English bilinguals in their English writing. It is evident that, except for the language-specific idioms and slogans, all the discursive conventions discussed above are also found in native varieties of English. The claim is simply that these discursive conventions are more typical or characteristic of Chinese English, or more exactly, Chinese English used during the turbulent political movement, the Great Proletarian Cultural Revolution, than of any native variety of English. Of course, this claim presupposes that writing from the same genre in both native varieties of English and Chinese English is used for the purpose of comparison.

4.3 Nativization and bilinguals' creativity

The analysis above shows that although English is used primarily as a performance language variety in China, the nativization process has already started. This lends support to the claim that a Chinese variety of English is emerging, be it labeled as 'Chinese English', 'China English', or 'Sinicized English'.¹² In other words, the development of English in the Chinese context is approaching the second phase.¹³ This echoes the argument in Kirkpatrick & Xu (2002). Provided that English is going to maintain its current status as an international language for an indefinite period of time and that China's current sociopolitical and linguistic policies are not discontinued, the use of English in China will be more and more common. This, in turn, foretells the greater degree to which English is going to be nativized. It is even possible that in the future English may become an institutionalized language variety in China. When this happens, English in the Chinese context reaches the third, and also the last, phase of nativization.

Through the investigation of the discursive nativization of English in China, this article also sheds light on the study of bilinguals' linguistic creativity, in this case Chinese-English bilinguals' creativity shown in their writing in English, their nonnative language. In a nutshell, these bilinguals utilized nativized discourse strategies to convey information about their unique culture, which is difficult, if not impossible, to be communicated through the use of a native variety of English. As this study shows, bilinguals' creativity is represented not only by the use of

¹² When debating on the issue whether there exists a legitimate Chinese variety of English, some scholars distinguish between China English (or Sinicized English) and Chinese English (or *Chinglish*). Whereas the former is considered to be a legitimate variety, which refers to English used by the Chinese in China, based on standard English, and with Chinese characteristics (e.g., Wang 1991), the latter is not. Instead, it usually denotes an interference variety. For detailed discussions about the distinction see, for instance, Jiang (1995) and Zhang (1997). In this paper, the terms 'Chinese English' and 'China English' are used interchangeably in the sense of Wang (1991).

¹³ B. Kachru (1992a:56) proposes that there are three phases that institutionalization of a nonnative variety of English experiences. The first phase is the 'non-recognition' of the local variety. The second phase is the 'development of varieties within a variety', when the local variety is used widely but is still socially stigmatized. In the third phase the local model becomes the norm and is consequently socially acceptable.

'culture words' (e.g., Xu 1987), but also by the innovative use of speech acts, rhetorical devices, idioms, or slogans that are characteristic of the native culture. Although linguistic nativizations that result from such creativity may not be readily recognized or accepted by native speakers of English, they are an indispensable means of communicating the unique Chinese culture to the outside world and, therefore, should be accepted and even encouraged.

5. Pedagogical implications

As this study indicates, English as a language of wider communication has 'become pluricentric, and carries the weight of various sociocultural identities' (Y. Kachru 2001:350). In other words, as in China, the English language has been and is being nativized in all sociocultural contexts where it is assumed that native speakers are a minority, if there are any at all (e.g., Bolton 2001, Forthcoming; D'souza 2001; Y. Kachru 1998). As Y. Kachru (2001) rightly points out, this has brought new questions and challenges to the profession of English teaching. For instance, one of the crucial questions that must be asked is which model of English is the most appropriate for teaching.

It is a fact that Chinese people use English not only with native speakers, but also with those from countries where English is either a foreign language, for example, Japan, the Republic of Korea, and Russia, or a second language, for example, India, Malaysia, and Singapore (e.g., Gao 2001, Pang, Zhou, & Fu 2002). Furthermore, as is claimed by Kirkpatrick & Xu (2002), in China, as in many other Asian countries, people use English with predominantly nonnative speakers of English, who now outnumber native speakers of English.

In light of this situation, in the teaching of English, it may not be advisable for China to adopt invariably the established norm(s) of English in the Inner Circle countries, for example, BBC English, or General American English. Instead, it is more sensible to determine the teaching model on the basis of the interlocutors with whom English is used. In other words, 'the function to which English is to be put is ... crucial in determining which variety is the most appropriate' (Kirkpatrick & Xu 2002:275). Consequently, students in China must be given the opportunity to be exposed to all major varieties of English in the world — native and nonnative varieties as well. This actually echoes the overall sentiment of most Chinese college students, according to Kirkpatrick & Xu (2002).

When discussing the teaching of World Englishes, B. Kachru (1992a) identifies six different fallacies concerning the users and uses of English. This provides strong theoretical support for the adoption of functionally oriented teaching models.¹⁴ It is understandable that some people may be concerned with the intelligibility problem if the adoption of local models is encouraged in the teaching of English. However, as researchers argue (e.g., Nelson 1992, 1984, 2001; Smith 1992;

¹⁴ Three relevant fallacies out of the six are: 1) in the Outer Circle and the Expanding Circle, people learn English essentially to interact with its native speakers; 2) people learn English as a tool to understand and teach what is generally termed the Judeo-Christian traditions; and 3) the goal of learning and teaching English is to adopt the native models of English.

Smith & Nelson 1985), the unintelligibility problem in World Englishes is mainly caused by the lack of exposure and sensitivity to the variety of English involved. This further helps to support the argument that invariably following the model established in the Inner Circle countries may not be the best policy.

6. Conclusion

This paper examines primarily the discursive nativization of English used in the Chinese context. In so doing it supports the argument that the Chinese variety of English is emerging, which in itself manifests Chinese-English bilinguals' linguistic creativity. Adopting a function-oriented standpoint, this paper also briefly discusses the issue of which model(s) of English should be taught in Chinese schools. Research on the nativization of English investigates this linguistic phenomenon that takes place mostly in the Outer Circle countries, such as India, Malaysia, and Singapore, whereas the nativization of English in the Expanding Circle countries is relatively inadequately documented.¹⁵ By investigating the discursive nativization of English used in China, a country in the Expanding Circle, this study aims at making some contribution to research on World Englishes.

REFERENCES

- ANDERSON, Beth, Benjamin FORTSON, Steven KLEINEDLER, & Hanna SCHONTHAL (eds.). 2001. *The American Heritage Dictionary*. New York: Dell Publishing.
- BOLTON, Kingsley. 2000. Language and hybridization: pidgin tales from the China coast. *Interventions* 2:1.35-52.
- . (ed.). 2001. Special issue on 'Hong Kong English: Autonomy and Creativity'. *World Englishes* 19:3.
- . 2002. Chinese Englishes: from Canton jargon to global English. *World Englishes* 21:2.181-99.
- . [Forthcoming]. *Chinese Englishes: A Sociolinguistic History*. Cambridge, UK: Cambridge University Press.
- BOLTON, Kingsley, & Q. S. TONG (eds.). 2002. Special issue on 'Interdisciplinary Perspectives on English in China'. *World Englishes* 21:2.
- BOLTON, Whitney, & David CRYSTAL (eds.). 1993. *The English Language*. (*The Penguin History of Literature*, volume 10). London: Penguin Books.
- BROWN, Adam (ed.). 2000. *English in Southeast Asia 99*. Nanyang Technological University, Singapore: National Institute of Education.
- BRUTT-GRIFFLER, Janina. 2002. *World English: A Study of Its Development*. Toronto: Multilingual Matters Ltd.

¹⁵ This, of course, is not to say that there has been no research on English used in the Expanding Circle countries. For example, *World Englishes* 21(2), which was guest edited by Bolton & Tong (2002), is devoted to the investigation into the sociolinguistics of English in the Chinese context. This special issue examines English in China from interdisciplinary perspectives, among them literary, educational, and linguistic.

- BURCHFIELD, Robert (ed.). 1994. *English in Britain and Overseas: Origins and Developments*. (The Cambridge History of the English Language, volume 5). Cambridge, UK: Cambridge University Press.
- CHENG, Chin-Chuan. 1992. Chinese varieties of English. In: B. Kachru (ed.), 162-77.
- CRYSTAL, David. 1985. How many millions? The statistics of English today. *English Today* 1.7-9.
- . 1997. *English as a Global Language*. Cambridge, UK: Cambridge University Press.
- DI, Ganzhi. 2001. *Yin-yang* dualism reflected in the ESL writing by Chinese students. Paper presented at the annual meeting of the International Linguistic Association. New York University, NY.
- D'SOUZA, Jean. 2001. Indian English. In: Thumboo (ed.), 3-17.
- DZAU, Y. F. 1990. *English in China*. Hong Kong: API Press Ltd.
- GAO, Liwei. 2001. The lexical acculturation of English in the Chinese context. *Studies in the Linguistic Sciences* 31:2.73-88.
- GOEBL, Hans, Peter NELDE, Zdenek STARY, & Wolfgang WOLCK (eds.). *Contact Linguistics*. Berlin: Walter de Gruyter.
- GUMPERZ, John (ed.). 1982. *Language and Social Identity*. Cambridge, UK: Cambridge University Press.
- HUNG, Tony T. 2002. English as a global language and the issue of international intelligibility. *Asian Englishes* 5:1.4-17.
- JIA, Yuxin, & Cheng CHENG. 2002. Indirectness in Chinese English writing. *Asian Englishes* 5:1.64-74.
- JIANG, Yajun. 1995. Chinglish and China English. *English Today* 11:1.51-3.
- KACHRU, Braj B. 1965. The Indianness in Indian English. *Word* 21.391-410.
- . 1980. Bilingualism. *Annual Review of Applied Linguistics* 1.2-18.
- . 1983. *The Indianization of English: the English Language in India*. New Delhi: Oxford University Press.
- . 1985. Standards, codification and sociolinguistic realism: the English language in the outer circle. In: Quirk & Widdowson (eds.), 11-30.
- . 1986a. The bilingual's creativity and contact literatures. In: B. Kachru, 1990:159-173.
- . 1986b. The bilinguals' creativity. *Annual Review of Applied Linguistics* 6.20-33.
- . 1987a. The bilingual's creativity: discursual and stylistic strategies in contact literatures. In: Smith (ed.), 125-40.
- . 1987b. The passed and prejudice: toward de-mythologizing the English canon. In: Steel & Threadgold (eds.), 245-56.
- . 1990. *The Alchemy of English: The Spread, Functions, and Models of Non-native Englishes*, 2nd ed. Urbana, IL: University of Illinois Press.
- (ed.). 1992a. *The Other Tongue: English Across Cultures*, 2nd ed. Urbana, IL: University of Illinois Press.
- . 1992b. The second diaspora of English. In: Machan & Scott (eds.), 230-52.
- . 1994. English in South Asia. In: Burchfield (ed.), 497-553.

- . 1996. English as a lingua franca. In: Gobel, Nelde, Stary, & Wolck (eds.), 906-13.
- KACHRU, Yamuna. 1987. Cross-cultural texts, discourse strategies and discourse interpretation. In: Smith (ed.), 87-100.
- . 1995. Contrastive rhetoric in World Englishes. *English Today* 41:21-31.
- . 1998. Culture and speech acts: evidence from Indian and Singaporean English. *Studies in the Linguistic Sciences* 28:1.79-98.
- . 2001. Discourse competence in World Englishes. In: Thumboo (ed.), 341-55.
- KAPLAN, Robert. 1966. Cultural thought patterns in intercultural education. *Language Learning* 16.1-20.
- KIRKPATRICK, Andy. 1991. Information sequencing in Mandarin in letters of request. *Anthropological Linguistics* 33:2.1-20.
- . 1995. Chinese rhetoric: methods of argument. *Multilingua* 14:3.271-95.
- KIRKPATRICK, Andy, & Zhichang XU. 2002. Chinese pragmatic norms and 'China English'. *World Englishes* 21:2.269-79.
- MACHAN, Tim, & Charles SCOTT (eds.). 1992. *English in Its Social contexts: Essays in Historical Sociolinguistics*. New York: Oxford University Press.
- MCARTHUR, Tom. 1987. The English languages? *English Today* 3:3.9-11.
- (ed.). 1992. *The Oxford Companion to the English Language*. Oxford, UK: Oxford University Press.
- . 1993. The English language or the English languages? In: Bolton & Crystal (eds.), 323-41.
- MOHAN, Bernard, & Winnie Au-Yeung LO. 1985. Academic writing and Chinese students' transfer and developmental factors. *TESOL Quarterly* 16.515-34.
- MYERS, James. 2000. Rhetorical style: Chinese encounter Aristotle. In: Brown (ed.), 71-82.
- NELSON, Cecil L. 1984. Intelligibility: the case of non-native varieties of English. University of Illinois at Urbana-Champaign, Ph.D. dissertation in Linguistics.
- . 1992. Intelligibility and non-native varieties of English. In: B. Kachru (ed.), 58-73.
- . 2001. Creativity and intelligibility in world Englishes. In: Thumboo (ed.), 35-44.
- PAIKEDAY, Thomas. 1985. *The Native Speaker Is Dead*. Toronto: Paikeday Publishing.
- PAKIR, Anne. 2001. English as a cross-cultural lingua franca: multiforms, multimedia, multidisciplinary. In: Thumboo (ed.), 77-90.
- PANG, Jixian, Xing ZHOU, & Zheng FU. 2002. English for international trade: China enters the WTO. *World Englishes* 21:2.201-16.
- PRIDE, John, & Ru-shan LIU. 1988. Some aspects of the spread of English in China since 1949. *International Journal of the Sociology of Language* 74.41-70.
- QUIRK, Randolph, & Henry G. WIDDOWSON (eds.). 1985. *English in the World: Teaching and Learning the Language and Literatures*. Cambridge, UK: Cambridge University Press.

- SMITH, Larry (ed.). 1987. *Discourse Across Cultures: Strategies in World Englishes*. London: Prentice-Hall International.
- . 1992. Spread of English and issues of intelligibility. In: B. Kachru (ed.), 75-90.
- SMITH, Larry, & Cecil L. NELSON. 1985. International intelligibility of English: directions and resources. *World Englishes* 4:3.333-42.
- STEEL, ROSS, & Terry THREADGOLD (eds.). 1987. *Linguistic Topics: Papers in Honor of M. A. K. Halliday*. Philadelphia: John Benjamins.
- TANG, Paul C. L. 1997. On the special logic thesis in Chinese philosophy. *Meta-philosophy* 28.371-84.
- THUMBOO, Edwin (ed.). 2001. *The Three Circles of English*. The Center for the Arts, National University of Singapore: UniPress.
- WANG, Rongpei. 1991. *Zhongguo yingyu shi keguan cunzai* [China English is real]. *Jiefangjun Waiyu Xueyuan Xuebao* [Journal of the PLA Institute of Foreign Languages] 1:1.12-8.
- XU, Guozhang. 1987. Code and transmission in cross-cultural discourse: a study of some samples from Chinese and English. In: Smith (ed.), 66-73.
- YOUNG, Linda W. 1982. Inscrutability revisited. In: Gumperz (ed.), 72-84.
- ZHANG, Ailing. 1997. China English and Chinese English. *English Today* 52.39-41.
- ZHANG, Hang. 2002. Bilinguals' creativity in Chinese English: Ha Jin's *In the Pond*. *World Englishes* 21:2.305-15.
- ZHAO, Yong, & Keith CAMPBELL. 1995. English in China. *World Englishes* 14:3.377-90.
- ZHOU, Zhipai, & Wenchi FENG. 1987. The two faces of English in China: Englishization of Chinese and nativization of English. *World Englishes* 6:2.111-25.
- ZUO, Biao. 2001. Lines and circles, West and East. *English Today* 67.3-8.

THE ACCENTUAL SYSTEM OF MALLABIA BASQUE

José Ignacio Hualde*, Pello Mugarza†, and Koldo Zuazo‡

**University of Illinois at Urbana-Champaign*

j-hualde@uiuc.edu

†,‡*Euskal Herriko Unibertsitatea (UPV/EHU)*

fvpzuzek@vc-ehu-es

In this paper we examine accentuation in Mallabia Basque. This is an accentual system that is clearly transitional between the northern Bizkaian pitch-accent system (with a lexical distinction between accented and unaccented words, and phrase-final or phrase-penultimate unmarked accentuation) and the central Basque systems (with accent on the post-initial or, in the marked case, on the initial syllable). The study of the accentual patterns in Mallabia Basque offers us important cues to understand the historical connection between these two very different prosodic systems.

1. Introduction

In a broad western and central area of the Basque-speaking territory, we find accentual systems with regular post-initial accent (where some lexical exceptions may receive initial accent instead) (Txillardegi 1984; Hualde 1991, 1999a). In the northern part of Bizkaia, on the other hand, we find a rather different accentual system, reminiscent of that of Tokyo Japanese, where there is a lexical distinction between accented and unaccented words, the latter subject to a rule of phrase-final accentuation in certain contexts (Hualde 1988, 1991, 1999a; Elordieta 1997; cf. also Azkue 1931-32, Basterrechea 1974-75, Michelena 1972). In Hualde 2003 it is claimed that this Northern Bizkaian system represents a more conservative stage and several historical developments are proposed for the emergence of post-initial accent as a general pattern. In particular, it is proposed that the accentual system employed nowadays in the Mallabia dialect is a transitional type that sheds light on one of these possible diachronic routes. In this paper we describe the patterns of accentuation in Mallabia Basque in greater detail, since it demonstrates a particularly interesting prosodic development and since there are no other published descriptions of this accentual system beyond the very brief description offered in Hualde (2003).¹

Mallabia is a township of approximately 1110 inhabitants located about 15 km south of Markina and only 3 km to the west of the urban area of Ermua and Eibar. Its accentual system, although related to those of the Markina and Gernika areas, has a number of very interesting specific features. As mentioned, it represents a transitional type from the Northern Bizkaian pitch-accent system towards

¹ Although Gaminde's (1998) survey of Basque accentual types contains a brief mention of the Mallabia dialect (pp. 83-4), it overlooks the most interesting features of this system.

the system with generalized accent on the second syllable (with exceptional accent on the first) found in some other neighboring areas such as Durango and Elorrio.

2. Accentual patterns

In their citation form (which for nouns and adjectives is the absolutive singular, viz., stem + article /a/ — or /e/ after a high vowel), words in Mallabia Basque can be classified along two accentual dimensions: the position of the accent and the shape of the accentual pitch contour. Considering pitch-accent shape first, in some words, the pitch rises on the accented syllable and has an abrupt fall on the post-tonic. We will refer to this pitch-contour as 'acute accent'. In a lexically contrastive fashion, other words bear a 'grave accent' instead, in which the pitch also rises on the accented syllable but falls gradually towards the end of the word or phrase. Regarding the position of the accent, on the other hand, the accent may be located either on the first or on the second syllable of the word but not on any other syllable. All words accented on the initial syllable have an acute accent, whereas among those accented on the second, some have acute accent and some others have grave accent. By far the largest group is that of words with grave accent (on the second syllable) in their citation form. All of this is summarized and illustrated with examples in (1).²

(1) Words in citation form

I. Marked classes: Acute accent (rapid fall)

Ia. Acute initial accent: *égixe* 'truth'

Ib. Acute post-initial accent: *errótarrixe* 'whetstone'

II. Unmarked class: Grave accent (slow fall). Always on the second syllable: *mendixe* 'mountain', *erròtarixe* 'miller'

Minimal or near-minimal pairs can be found across both accentual dimensions of contrast, as shown in (2):

(2) (Near-)minimal pairs

a. Acute post-initial vs. grave

frutérie 'fruit-bowl' vs. *frutèrie* 'fruit-seller'

atákie 'attack' vs. *atàkie* 'gate'

sagútxue 'little mouse' vs. *sagùtxue* 'shrew'

errótarrixe 'whetstone' vs. *erròtarixe* 'miller'

b. Acute initial vs. acute post-initial

áskurie 'itch' vs. *askórie* 'axe'

c. Acute initial vs. grave post-initial

ártie 'art' vs. *artie* 'live-oak'

bárrie 'bar' vs. *barrie* 'laughter'

básie 'base' vs. *basie* 'mud'

párie 'pair' vs. *parie* 'shovel'

² All examples in this paper are cited using Basque orthographic conventions: *x* is a voiceless prepalatal fricative, *tx* the corresponding affricate, *tz* a voiceless dental fricative, *ll* a palatal lateral, *tt* a voiceless palatal stop, *r* a rhotic tap between vowels and *rr* an alveolar trill.

sárie 'large basket'	vs.	sarie 'net'
básue 'drinking glass'	vs.	basùe 'forest'
págue 'payment'	vs.	pagùe 'beech tree'
sérué 'zero'	vs.	serùe 'sky'
sáltzie 'the selling'	vs.	saltzie 'sauce'
bégire 'looking'	vs.	begìre 'more or less'

The acute/grave distinction is illustrated with a minimal pair in Figure 1.

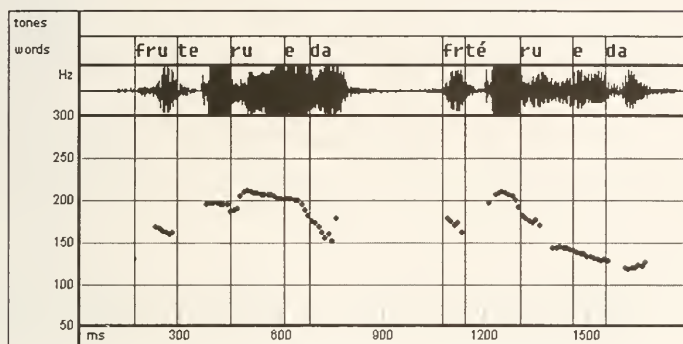


Figure 1: (a) *frutèru e da* 'it is the fruit seller'
(b) *frutéru e da* 'it is the fruit bowl'

As we can see in Figure 1, in both examples the pitch rises on the second syllable (the curve is interrupted here because of the voicelessness of the /t/). The difference is that in Figure 1a (lefthand example) the pitch continues rising in the post-tonic before it starts to fall. In Figure 1b (righthand example), on the other hand, there is a rise and a fall within the accented syllable and a much steeper fall to the end of the utterance.

If we consider the behavior of these different accentual classes in non-phrase-final position, we discover that acute-accent words have a rise-fall pitch accent (as in their citation form), which causes the downstep of the accent on a following word. On the other hand, grave-accent words show a relatively high, slowly falling plateau after the rise and, crucially, do not induce downstep of the accent on the next word in the phrase. Thus, if we compare, for instance, the examples in (3), a clear difference is that in the second example of each pair of sentences the second accent is downstepped (reduced) with respect to the first accent, whereas in the first example of each pair either the second peak is higher or both accentual peaks reach approximately the same level:

- (3) 1. lagùnan alabie da 'it is the friend's (sg) daughter'
lagúnen alabie da 'it is the friends' (pl) daughter'
2. lagùnan alábak die 'they are the friend's (sg) daughters'
lagúnen alábak die 'they are the friends' (pl) daughters'

In Figure 2 and Figure 3, we include F_0 tracings for the two pairs in (3).

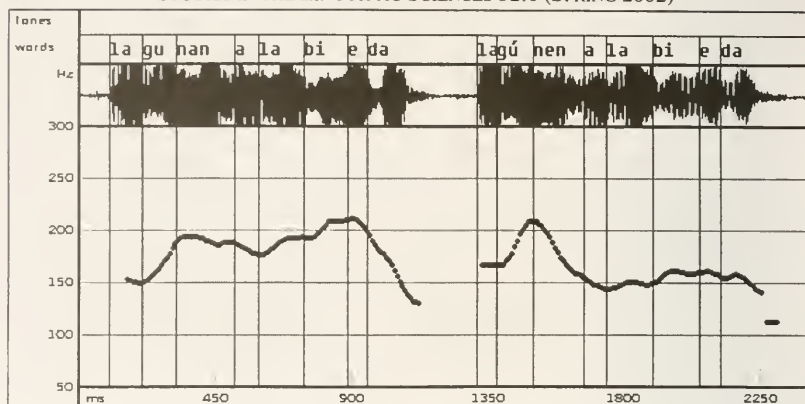


Figure 2: (a) *lagùnan alàbie da* 'it is the friend's (sg) daughter'
(b) *lagúnen alàbie da* 'it is the friends' (pl) daughter'

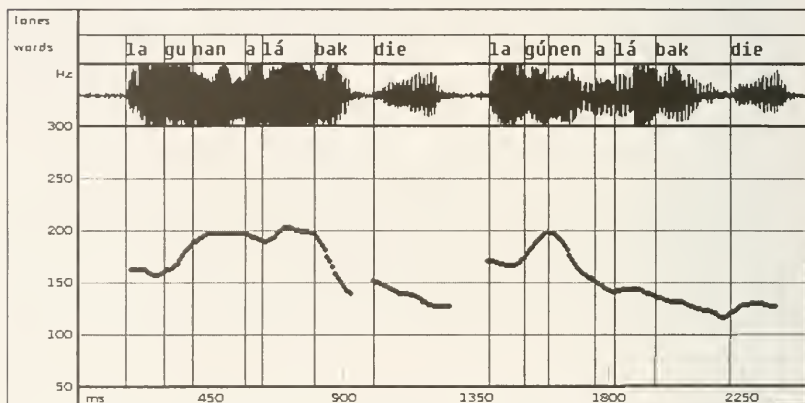


Figure 3: (a) *lagùnan alábak die* 'they are the friend's (sg) daughters'
(b) *lagúnen alábak die* 'they are the friends' (pl) daughters'

As can be seen in Figure 2 and Figure 3, a major difference between the (a) and (b) contours is in the scaling of the second peak. Both *alàbie* 'the daughter' in Figure 2 and *alábak* 'the daughters' in Figure 3 present a much lower peak in the (b) example than in the (a) example. The acute accent of the genitive plural form *lagúnen* 'of the friends' is manifested as a steep fall after the accented syllable which triggers the reduction of the second accent in the phrase, whether this second accent is grave, as in Figure 2b, or acute, as in Figure 3b. On the other hand, in the (a) examples, where the genitive singular *lagùnan* 'of the friend' has a grave accent, there is no downstep. The second accent of the phrase is realized as a higher peak than the first accent in both Figure 2a and Figure 3a.

Clearly the acute/grave distinction in Mallabia is related to the accented/unaccented distinction found in other Bizkaian varieties (see Hualde 1991,

1996, 1997, 1999a, 2000; Hualde & Bilbao 1992, 1993; Hualde, Elordieta & Elordieta 1994; Elordieta 1997; Hualde, Elordieta, Gaminde & Smiljanic 2002). The main difference is that instead of the high plateau or slightly rising contour up to the end of the phrase that, with lexically unaccented words, we find in other Northern Bizkaian varieties, in Mallabia we find a slowly falling contour beginning in the post-tonic. This has resulted in a reinterpretation in the perception of prominence, as argued in Hualde 2003.³

Accentual distinctions play a major role in inflected nominal forms in Basque. For words in the unmarked class, singular and plural forms often have different accentuation. In the next section we examine the accentuation of inflected nominals.

3. Accentuation of inflected nominal forms

As in other western and central Basque varieties (many Bizkaian and Gipuzkoan, and some High Navarrese varieties), singular and plural forms often contrast in accentuation. An interesting fact of the Mallabia dialect, not found in any other variety that has been described until now, is that words with vowel-final bisyllabic stems differ from other words in their accentuation in the plural. Furthermore, /i/-final stems, which undergo epenthesis of a voiceless prepalatal fricative *x* /*ʃ*/ between the stem and vowel-initial inflectional suffixes, are treated as consonant-final for accentual purposes.

The unmarked pattern is grave accent in most singular forms and post-initial acute accent in the plural and ablative singular, as shown in the following partial paradigm for /*mendi*/ 'mountain'.⁴

(4) *mendi* 'mountain'

	sg	pl
Absolute	<i>mendixe</i>	<i>mendíxek</i>
Dative	<i>mendixai</i>	<i>mendíxei</i>
Genitive ⁵	<i>mendixan</i> N	<i>mendíxen</i> N
Benefactive	<i>mendixantzat</i>	<i>mendíxentzat</i>
Inessive	<i>mendixen</i>	<i>mendíxetan</i>
Allative	<i>mendire</i>	<i>mendíxeta(ra)</i>
Comitative	<i>mendixakin</i>	<i>mendíxekin</i>
Ablative	<i>mendítitik</i>	<i>mendíxetatik</i>

By and large, the accentuation of these forms corresponds closely to what has been described for some northern Bizkaian varieties (the Gernika-Getxo type, Hualde 1999a, among others). For the most part, the grave-accent forms correspond to the unaccented forms of these other varieties. In particular, the anomalous accentuation of the ablative singular (which has acute accent) with respect to

³ An intermediate step in this evolution undoubtedly was the retraction of the accent from the final syllable of the phrase to the penultimate, as in neighboring Markina and Ondarroa (Hualde 2000, 2003).

⁴ *mendi* is the bare stem. The citation form is *mendixe*. Bare stems can be used without inflectional suffixes if accompanied by modifiers; e.g., *lau mendi* 'four mountains'.

⁵ The genitive must be followed by a noun; e.g. *mendixan isèna* 'the mountain's name' or by a determiner; e.g. *mendixana* 'the one of the mountain' vs. *mendíxena* 'the one of the mountains'.

other singular forms is general in Bizkaian. One difference with respect to other dialects is found in the comitative, which is accented in both singular and plural in other Bizkaian varieties, although the accent oftentimes falls on different syllables in the two numbers, e.g.: Getxo sg *laguná(g)as* 'with the friend', pl *lagúnakas* 'with the friends'; Gernika sg *lagunégas* 'with the friend', pl *lagúnekin* 'with the friends'; Markina sg *lagúnas*, pl *lagúnekin*.

We find the same pattern with all consonant-final and (epenthesis-triggering) /i/-final stems and also with vowel-final stems of three or more syllables, as illustrated in (5) with the absolutive singular and plural:

- (5) a. Consonant-final stems (including /i/-final)
- | | | |
|--------------------------|-------------|--------------|
| uninflected stem | sg | pl |
| ogi 'bread' | ogixe | ogíxek |
| gison 'man' | gisóna | gisónak |
| lagun 'friend' | lagúne | lagúnek |
| katedral 'cathedral' | katèdrala | katédralak |
| katamixer 'squirrel' | katàmixerra | katámixerrak |
| bersolari 'verse-singer' | bersòlarixe | bersòlarixek |
- b. Vowel-final stems of three or more syllables
- | | | |
|--------------------------------|------------------|------------------|
| alaba 'daughter' | alàbie | alábak |
| abade 'priest' | abàdie | abádiek |
| kategorixa 'category' | katègorixie | katégorixak |
| astakume 'baby donkey' | astàkumie | astákumak |
| basamortu 'desert' | basàmortue | basámortuek |
| arrijasotzaille 'stone-lifter' | arrijasotzaillie | arrijasotzaillak |

On the other hand, bisyllabic stems ending in a vowel other than /i/ show a shift of the accent to the initial syllable in all plural forms (but not in the ablative singular, even though this form also has an acute accent), this is illustrated in (6) with a partial paradigm of *etxe* 'house'. Further examples, in the absolutive, are given in (7):

- (6) Vowel-final bisyllabic stems: *etxe* 'house'

	sg	pl
Absolutive	etxìe	étxiek
Ergative	etxìek	étxiek
Dative	etxìai	étxiei
Genitive	etxìan N	étxien N
Benefactive	etxìantzat	étxientzat
Inessive	etxien	étxietan
Allative	etxèra	étxieta(ra)
Ablative	etxétik	étxietatik

- (7) Other examples of vowel-final bisyllabic stems

lora 'flower'	lorìe	lórak
kale 'street'	kaliè	káliek
asto 'donkey'	astùe	ástuek
esku 'hand'	eskùe	éskuek

Phrase-final accent is generally avoided. This results in retraction of the accent to the first syllable of two-syllable words in final position:

(8) Phrase-final retraction

baltz 'black'*bàltza* 'the black one'*baltzà da* 'it is the black one'cf. pl: *bàltzak die* 'they are black', *bàltzak* 'the black ones'

The phrase-final retraction rule has an exception in allative forms, which can be optionally contracted: *etxèra* ~ *etxà* 'to the house. allative sg.'

What we have said so far applies only to words in the unmarked class (=grave accent in citation form, abs sg). All words that have an acute accent in the absolutive singular (and in their uninflected form, whether on the initial or on the post-initial syllable), keep the accent on that syllable throughout their inflectional paradigm. With these exceptional words there is no accentual distinction between singular and plural:⁶

(9) Marked class: acute stems

	sg	pl
músker 'lizard'	múskerra	múskerrak
txístu 'flute'	txístue	txístuek
léngusu 'cousin'	léngusue	léngusuek
báserri 'farm'	báserrixe	báserrixek
arráno 'eagle'	arránue	arránuek
basérrittar 'farmer'	basérrittarra	basérrittarrak

Consequently with what we have said, many of the minimal pairs given above in (2) are not distinguished in the plural:

(10) Minimal pairs

fruteru 'fruit-seller'	frutèrue	frutéruék
frutéru 'fruit-bowl'	frutéruék	frutéruék

4. Marked words (acute accent in citation form)

As mentioned, words with an acute accent in their citation form are exceptional (i.e., they form a relatively small class). As in other western and central Basque dialects with an accentually exceptional or marked class, these words fall into two main groups: (a) compounds and words with certain derivational suffixes and (b) borrowings, some of them very old (Michelena 1972, Hualde 1992, among others). In addition, there is a small residue of words of unclear origin. In general, it seems that the Mallabia dialect has fewer marked words than some other Bizkaian varieties whose lexicon has been explored from this point of view, such as Lekeitio (Hualde, Elordieta & Elordieta 1994) and Getxo (Hualde & Bilbao 1992).

⁶ This is also true of items in the marked accentual class in other western and central dialects, as already noticed in Michelena (1977).

4.1. Compounds

Many compounds have acute accent, including productively created compounds. The accent may be on the first or the second syllable. Most compounds with initial accent have a monosyllabic first member (i.e., the accent tends to occur on the first member of the compound). Nevertheless, it is not the case that all compounds have marked accent. Instead, we find some rather idiosyncratic subregularities. For instance, compounds with *-ar* 'male' and *-eme* 'female' have acute accent, but those with *-kume* 'child, offspring' do not:

(11) Compounds

Ia. Acute initial accent

óill-arra 'rooster' (< oilo 'chicken'), ást-arra 'male donkey' (< asto 'donkey'), ást-emie 'female donkey', kát-arra 'tom cat' (< katu 'cat'), kát-emie 'she-cat', áittitte 'grandfather' (< aita 'father'), ánk-utzik 'with bare legs' (< (h)anka 'leg'), árañ-eun 'day before yesterday' (standard Basque heren-egun, lit. 'third day'), báltz-unie 'black spot', bás-errixe 'farm' (< baso 'forest' + (h)erri 'village'), és-ola 'pole' (< hesi 'fence' + ola 'board'), étx-aurrice 'façade' (< etxe 'house' + aurre 'front'), il-gora 'waxing of the moon' (lit. 'moon-up'), lén-gosue 'first cousin', ít-aurren 'guiding oxen' (< idi 'ox' + aurre-an 'in front of').

Ib. Acute post-initial accent

abáde-gaixe 'seminarian' (cf. abàdie 'priest'), adár-soñue 'joke' (lit. 'horn-play'), afál-ostie 'after dinner' (cf. afàrie 'supper'), aitté-besuetako 'godfather' (< aita 'father' + beso-etako 'of the arms'), aitté-ñarreba 'father-in-law', aitté-ordie 'step-father', aitté-santue 'the Holy Father', alb-ágiñe 'side tooth', amá-iru 'thirteen', amá-saspi 'seventeen', amá-sulo 'mommy's child' (lit. 'mother-hole'), anáixa-ordie 'step-brother', anká-lusie 'long legged' ank-ártie 'between the legs', ank-ókerra 's.o. with crooked legs', aó-sabala 'talkative' (lit. 'wide-mouth'), arró-putze 'proud', artá-berdie 'green corn', artá-lorie 'corn flower', asá-burue 'head of cabbage', bet-ókerra 'with crooked eyes' (bet-, compositional form of begi 'eye' + oker 'twisted'), asté-lena 'Monday' (< aste 'week' + lehen 'first') (BUT astègune 'week day'), barré-gurie 'laughter' (< barre 'laugh' + gura 'desire'), jakín-gurie 'desire to know', esán-gurie 'meaning', bas-éri-ttara 'farmer', baskál-ostie 'after dinner', burú-aundixe 'big-head, type of bird, lanius collurio' (BUT burú-bidie 'reasoning'), burú-surixe 'white headed', burdúntzixe 'roasting spit' (< burdin 'iron' + untzi 'pot'), dirú-aldixe 'money-time, time when there is money', egú-astena 'Wednesday', egú-barrixe 'Christmas' (lit. 'day-new'), egú-éna 'Thursday', egún-sentixe 'dawn', egúr-iketza 'wood coal', egón-esiñe 'restlessness', biótz-errie 'heart burn', erbí-txakurre 'hare dog, hound', errú-bakue 'blameless', esné-beixe 'milk cow', estú-aldixe 'hard times' (lit. 'narrow-time'), gabón-

sar 'New Year's eve', geldí-unie 'resting time', gibél-urdiñe 'type of mushroom, *russula viriscens*' (lit. 'back-blue'),⁷ ipúrt-argixe 'lighting bug' (lit. 'buttocks-light'), ján-arixe 'food', negár-gurie 'desire to cry', jun-étorrixie 'round trip' (lit. 'go-come') kardú-latza 'hard thistle' (BUT kardù-berie 'soft thistle'), karrámarrue 'crab', lurr-íkerie 'earthquake', otz-íkerie 'tremor', mutíl-sarra 'bachelor' (lit. 'boy-old'), udé-barrixe 'spring' (lit. 'summer-new').

II. Grave accent

astà-kumie 'baby donkey', katù-kumie 'kitten', axè-lekue 'windy spot', basà-katue 'wild cat', basò-mutille 'lumberjack' (lit. 'forest-boy'), bisì-modue 'lifestyle', eskòn-barrixe 'newly-wed', jaix-ègune 'holiday', jan-èdana 'food and drink'.

4.2. Derived words

Most derived words have unmarked grave accent, but some derivational suffixes induce acute accent. Again, the grave vs. acute distinction corresponds directly to the unaccented/accented contrast of other dialects and, by and large, the same words and classes of words are found in the two groups as in other dialects.

(12) Derived words

I. Acute accent

-ti: áis-tixe 'oak grove', sagás-tixe 'appletree grove', maás-tixe 'vineyard', ló-tixe 'sleepy'

-ka: adár-ka 'hitting with horns' (cf. adarra 'horn'), arrí-ke 'throwing stones'

-kera: adár-kerie 'horn shape' (cf. adarkadie 'length of horns'), ibíł-kerie 'manner of walking', orrás-kerie 'way of combing, hairdo', lús-erie 'length'

-tza: dirú-tzie 'amount of money', lastó-tzie 'pile of straw', lagún-tzie 'help' morrón-tzie 'serfhood'

-ero: áste-ro 'weekly', egún-ero 'daily', íll-ero 'monthly', úrte-ro 'yearly'

-dun: ardí-dune 'sheep owner', motór-dune 'motorized' (BUT ulesuridúne 's.o. with white hair')

-en: báltz-ena 'blackest'

-ago: báltz-ague 'blacker'

-egi: báltz-egixe 'too black'

II. Grave accent

-garri: berò-garrixe 'heater'

-tasun: berò-tasune 'heat'

-lari: bersò-larixe 'verse-singer'

⁷ As remarked in Michelena (1970) (see also Trask 1997: 267), this example shows that the semantic field of *urdiñ*, nowadays 'blue' or 'gray', must have formerly included the meaning 'green' (in which sense it has been replaced by the borrowing *berde*), since this mushroom's color is a deep green.

4.3. Loanwords

Many words in the acute accent class are borrowings, both very old and recent. Whereas some of them preserve the etymological place of the accent, many others do not:

(13) Acute-accent borrowings

Ia. Initial

ábixie 'nest' < Lat. *cavea*, ántzarra 'goose' < Lat. *anser*, árbolie 'tree', árrixkue 'risk', átuen 'immediately' < Sp. *en el acto*, bésperie 'night before' < Sp. *víspera*, dántzie 'dance', déndie 'store', dénporie 'time' < Lat. *tempora*, dómekie 'Sunday' < Lat. *dominica*, dúrdurie 'type of bird, *turdus viscivorus*', gánbarie 'attic' < Lat. *camera*, gónie 'skirt', íngurue 'surrounding area' < Lat. *in gyrum*, káixolie 'cage' < Rom. < Lat. *caveola*, kíkerie 'pitcher' < Sp. *jícara*, kínddie < Sp. *guinda*, kórtie 'yard', kúlpie 'blame', kúbixe 'Guinea pig' < Sp. *cuy*, lándarie 'plant', lángie 'plank', lápitze 'pencil' < Sp. *lápiz*, lékue 'place' < Rom. *lueco*, líburue 'book', máixue 'teacher', mállue 'hammer', périxie 'market day' < Sp. *feria*, póspolue 'match' < Sp. *fósforo*, sápatue 'Saturday', téllatue 'roof' < Rom., cf. Sp. *tejado*.

Ib. Post-initial

arkóndarie 'shirt' < Old Sp. < Arab., askórie 'axe' < Lat. *asceola*, asúkrie 'sugar', bentánie 'window', domínistikue 'sneeze' < Lat. *dominus tecum*, erlójue 'watch', erréginie 'queen' < Lat. *regina*, errékie 'creek' < Rom., cf. Occitan *rec*, eskólie 'school', itxúrie 'aspect' < Sp. *hechura*, kittérrie 'guitar', koméntue 'convent', martítzena 'Tuesday' < Lat. *Martis eguna*, puxíkue 'bladder' < Lat. *uessica*, afróntue 'squall', bandérie 'flag'.

Many borrowings have an acute accent on a different syllable from the source:

(14) Borrowings with accent shift

abíonetie 'small plane' < Sp. *avioneta*, abánikue 'fan' < Sp. *abanico*, aiúntamentue 'town hall' < Sp. *ayuntamiento*, dinámittie 'dynamite' < Sp. *dinamita*, dispáratie 'nonsense' < Sp. *disparate*, elétronikie 'electronics' < *electrónica*, errébie 'magazine' < Sp. *revista*, errégalue 'gift' < Sp. *regalo*, errémedixue 'remedy' < Sp. *remedio*, errépublikie 'republic' < Sp. *república*, errétolíkue 'scolding' < Sp. *retórica*, eskármientue 'punishment' < Sp. *escarmiento*, eskúpetie 'shot gun' < Sp. *escopeta*, espáratrapue < Sp. *esparadrápo*, estémangue 'stomach' < Sp. *estómago*, gasóliñie 'gasoline' < Sp. *gasolina*, kalábásie 'pumpkin' < Sp. *calabaza* (BUT *kalidadie* 'quality'), kontrábandue 'smuggling' < Sp. *contrabando* (BUT *kontràbandistie* 'smuggler'), mansánillie 'camomille' < Sp. *manzanilla*.

As in other Bizkaian dialects, specific accentual patterns are applied to Spanish borrowings with certain endings. For instance, borrowings from Spanish *-ero* words have acute accent when the meaning is 'place, receptacle' but grave

accent when the meaning is 'profession, occupation' (Hualde 1999b), as we see in the example in Figure 1.

Although clearly there are certain tendencies, in general it appears that the accentuation of borrowings is not completely predictable. Thus, to give just one example, whereas Sp. *grámo*, *gráno* give *grámue*, *gránue*, respectively, from Sp. *grádo* we have *gradùe*.

4.4. Other exceptional items

Finally, besides compounds and derived words and various types of loanwords, there is also a relatively small group of acute-accent words of uncertain etymology. By and large, the same items belong to this group in all western and central varieties that recognize an accentually exceptional class:

(15) Words of uncertain etymology with acute accent

Ia. Initial

bélarrixé 'ear', ábaue 'honey comb', ádoríe 'vigor' (< Sp. ardor?), árrautzié 'egg', átzíe 'back part' áurrié 'front', báskarixé 'dinner', érdérie 'Spanish lg.', éuskeríe 'Basque lg.', éuskixé 'sun' (< eguzki < egu(n) 'day' + ?), béstíe 'other', gánoríe 'skill', gárrasixé 'cry', íntxaurre 'walnut' (< ? + ur 'hazelnut'), ístarríxé 'throat', ítaíxe 'scythe' (< igitai), máíxe 'table', mállíe 'level', múskerra 'lizard', órratza 'needle', símaurre 'fertilizer', tálué 'pancake', txálué 'applause', txíngorra 'hail', úsaba 'boss' (< ugazaba).

Ib. Post-initial

adábagixé 'mending', arránue 'eagle', asérixé 'fox', berákatzá 'garlic', eríñotzá 'laurel', errúkixé 'pity', goróstixé 'holly, ilex aquifolium', lapíkue 'cooking pot', orkátillíe 'ankle', otzárie 'basket', úrdaíxe 'lard, bacon' (< urde 'pig' + gai 'matter'?).

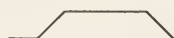
5. A comparative/diachronic perspective

The acute/grave contrast lends the Mallabia accentual a certain amount of complexity. As mentioned at the beginning of the paper, undoubtedly this contrast is diachronically related to the accented/unaccented distinction found in Northern Bizkaian, including Markina, just a few kilometers to the north of Mallabia.

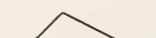
In an area from Getxo to Gernika, phrases composed entirely of lexically unaccented items receive final accent when focalized or when uttered in isolation.⁸ In these phrases, the pitch rises at the beginning of the phrase and stays high up to the last syllable, where it falls, as shown in the example on the left in (16) with a schematic contour. Accented words, on the other hand, present a sharp fall immediately after the accented syllable, as in the corresponding plural example on the right:

⁸ Unaccented words or phrases in isolation receive a pitch accent in Basque; just as in English or Spanish a word pronounced in isolation will regularly bear a pitch accent.

(16) Gernika-Getxo



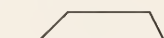
lagunena dá
'it is the one of the friend'



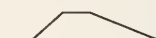
lagúнена da
'it is the one of the friends'

In Lekeitio we find essentially the same contrast, with the difference that (almost) all lexical accents have shifted to the penultimate syllable of the word:

(17) Lekeitio



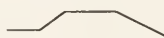
lagunena rá
'it is the one of the friend'



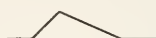
lagunéna ra
'it is the one of the friends'

In Ondarroa and Markina, unaccented phrases receive phrase-penultimate accent, instead of phrase-final accent. In Markina, where, in this particular case, the lexical accent of the plural form is on the same syllable as in Gernika, the resulting contrastive patterns are as in (18), whereas in Ondarroa, where the accent of the plural form has been shifted as in Lekeitio, the pitch patterns are as schematized in (19):

(18) Markina

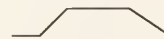


lagunaná da
'it is the one of the friend'



lagúнена da
'it is the one of the friends'

(19) Ondarroa



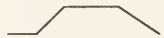
lagunaná ra
'it is the one of the friend'



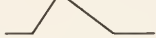
lagunána ra
'it is the one of the friends'

In all these systems, in the lexically unaccented case the pitch remains high up to the final (in Gernika-Getxo and Lekeitio) or the penultimate syllable (in Markina and Ondarroa) and this is the syllable that is perceived as carrying accentual prominence. In Mallabia, instead, we find a slow fall from the third syllable of the phrase to the end (generally more pronounced in the last two syllables), a fact that has caused the perception of prominence to be transferred to the second syllable:

(20) Mallabia



lagùnana da
'it is the one of the friend'



lagúнена da
'it is the one of the friends'

F₀ tracings for these Mallabia examples can be observed in Figure 4:

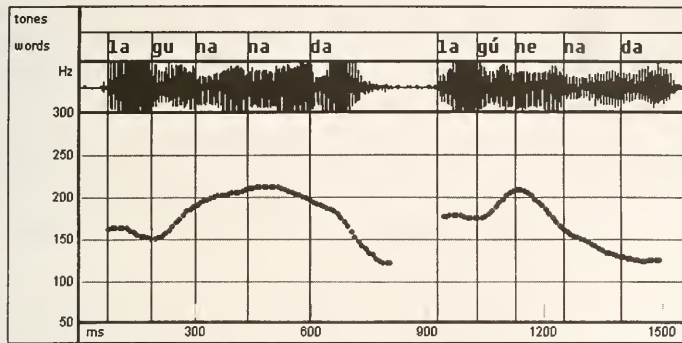


Figure 4: (a) *lagùnana da* 'it is the one of the friend'
(b) *lagúne na da* 'it is the one of the friends'

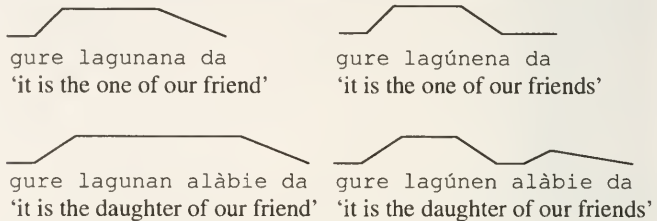
Notice that in the (a) example in Figure 4, after the abrupt rise on the second syllable, the peak continues rising and the fall does not begin until the penultimate syllable. There thus seems to be a certain amount of variation on the precise point where the fall starts in grave-accent words/phrases. Although the patterns do not differ greatly from those found in Markina, the location of the fall in unmarked words and phrases is not well-demarked with respect to the end of the phrase, and this seems to have led to a perceptual reanalysis where now the rise on the second syllable is seen as the locus of the accent.

It is easy to envision how a system of the Mallabia type can be transformed into a much simpler system. It would be enough for some children to fail to acquire the rather subtle acute/grave distinction. The result would be a system where most of the time the accent would be on the second syllable (< former grave and post-initial acute), and exceptionally we would have initial accent. In this reanalyzed system, thus, words accented on the second syllable, whether they used to bear a grave or an acute accent, would fall together in one class, and would contrast only with those words that are accented on the initial. This is indeed what we find in some neighboring areas such as Durango, Elorrio, and Elgoibar.

The accentual system employed by conservative Mallabia speakers (i.e., those with the acute/grave distinction), however, remains very similar to the Northern Bizkaian pitch-accent type, which has a contrast between lexically accented and unaccented words. In particular it is very close to the Markina subtype. In fact, whereas 'grave accent' is an appropriate characterization of the pattern that these words show in phrase-final position, we saw above that grave-accent words do not cause downstep of a following accent, unlike acute-accent words (Figure 2). This suggests that grave-accent words are actually lexically unaccented, and that the rise on the second syllable is a phrase-boundary phenomenon. This is confirmed when these words are placed phrase-medially, as in, *gure lagùnana da* 'it is the one of our friend', *gure lagúne na da* 'it is our friend's (sg) daughter'. In this medial context we can see that a grave-accent word such as *lagúne(a)* loses its prominence altogether (since the rise takes place on the second syllable of the

phrase), whereas this prominence is preserved in the corresponding acute-accent plural forms, conveyed by a rapid fall immediately following the accented syllable:⁹

(21) Mallabia: Grave vs. acute in non-phrase-final position



Consider the F_0 tracings in Figure 5:

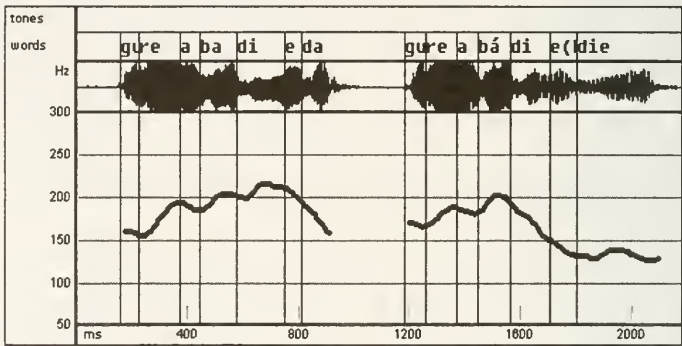


Figure 5: (a) *gure abàdie da* 'he is our priest'
 (b) *gure abádiek die* 'they are our priests'

In both examples in Figure 5, the main rise takes place on the second syllable of the phrase, *-re*. In the Figure 5a example, the rise continues until the antepenultimate syllable *-di-*, with small intervening dips that can be considered microprosodic effects produced by the voiced consonants /b/ and /d/ (realized as continuant in intervocalic position). In the Figure 5b example, containing the acute plural form *abádiek* 'the priests', the pitch rises up to the lexically accented syllable, where a steep fall starts.

Given these facts, we can analyze Mallabia Basque as possessing the same lexically accented vs. lexically unaccented distinction that we find in the Northern Bizkaian area. The main difference is in the tonal contour of phrases composed of unaccented items. In Lekeitio, for instance, these phrases clearly have final accent. In Ondarroa and Markina, they have penultimate accent in an equally clear fashion. Here is where we consistently find a fall in pitch. In Mallabia, on the other

⁹ Further research is, however, necessary to determine to what extent the initial rise has become a word-level, as opposed to phrasal, phenomenon in Mallabia.

hand, we have seen that the location of the fall is earlier than in these other dialects and appears to be more variable. This, we have suggested, has produced a shift of the locus of prominence to the second syllable of the phrase, characterized by a clear rise.

The other remarkable fact about Mallabia accentuation that we have noted is the leftwards shift of the accent in the plural of words with vowel-final bisyllabic stems.

APPENDIX 1: Map of Basque-speaking territory



APPENDIX 2: Mallabia and neighboring area



REFERENCES

- AZKUE, Resurrección María de. 1931-32. Del acento tónico vasco en algunos de sus dialectos. *Euskera* 4.282-318 & 6.3-50.
- BASTERRECHEA, José. 1974-75. Estudios sobre entonación vasca según el habla de Guernica. *Fontes Linguae Vasconum* 18.353-93 & 21.289-338.
- CHANG, Steve S., Lily LIAW, & Josef RUPPENHOFER. 1999. *Proceedings of the 25th Meeting of the Berkeley Linguistics Society. General Session and Parasession on loan word phenomena*. Berkeley, CA: Berkeley Linguistics Society.
- ELORDIETA, Gorka. 1997. Accent, tone and intonation in Lekeitio Basque. In: Martínez-Gil & Alfonso Morales-Front (eds.), 3-78. [Revised version in *Anuario del Seminario de Filología Vasca 'Julio de Urquijo'* 32.511-69.]
- FIKKERT, Paula, & Haike JACOBS (eds.). 2003. *Development in prosodic systems*. Berlin: Mouton de Gruyter.
- GAMINDE, Iñaki. 1998. *Euskaldunen azentuak*. Bilbao: Labayru Ikastegia.
- GUSSENHOVEN, Carlos, & Natasha WARNER (eds.). 2002. *Laboratory Phonology 7*. Berlin: Mouton de Gruyter.
- HUALDE, José I. 1988. A theory of pitch accent with special attention to Basque. *Euskara Biltzarra/Congreso de la Lengua Vasca* 1.53-60. II Euskal Mundu-Biltzarra/II Congreso Mundial Vasco. [Reprinted in *Anuario del Seminario de Filología Vasca 'Julio de Urquijo'* 22:3.915-19.]
- . 1991. *Basque phonology*. London: Routledge.
- . 1992. On the historical origin of Basque accentuation. *Diachronica* 10.13-50.
- . 1996. Accentuation and empty vowels in Ondarroa Basque: Against the concept of phonological derivation. *Lingua* 99.197-206.
- . 1997. *Euskararen azentuerak*. Donostia/San Sebastián & Bilbao: Gipuzkoako Foru Aldundia & Euskal Herriko Unibertsitatea (Supplements of *Anuario del Seminario de Filología Vasca 'Julio de Urquijo'* 42)
- . 1999a. Basque accentuation. In: van der Hulst (ed.), 947-93.
- . 1999b. Patterns of correspondence in the adaptation of Spanish borrowings in Basque. In: Chang, Liaw, & Ruppenhofer, 348-58.
- . 2000. On system-driven sound change: Accent shift in Markina Basque. *Lingua* 110.99-129.
- . 2003. From phrase-final to post-initial accent in western Basque. In: Fikkert & Jacobs (eds.), 249-81.
- HUALDE, José I., & X. BILBAO. 1992. A phonological study of the Basque dialect of Getxo. *Anuario del Seminario de Filología Vasca 'Julio de Urquijo'* 26.1-117.
- . 1993. The prosodic system of the Basque dialect of Getxo: A metrical analysis. *Linguistics* 31.59-85.
- HUALDE, José I., Gorka ELORDIETA, & Arantzazu ELORDIETA. 1994. *The Basque dialect of Lekeitio*. Bilbao & Donostia-San Sebastián: Univ. del País Vasco

- & Supplements of *Anuario del Seminario de Filología Vasca 'Julio de Urquijo'*.
- HUALDE, José I., Gorka ELORDIETA, Iñaki GAMINDE, & Rajka SMILJANIC. 2002. From pitch-accent to stress-accent in Basque. In: Gussenhoven & Warner (eds.), 547-84.
- HULST, Harry van der (ed.). 1999. *Word prosodic systems in the languages of Europe*. Berlin: Mouton de Gruyter
- MARTINEZ-GIL, Fernando, & Alfonso MORALES-FRONT (eds.). 1997. *Issues in the Phonology and morphology of the major Iberian languages*. Washington, D.C.: Georgetown Univ. Press
- MICHELENA, Luis. 1970. Nombre y verbo en la etimología vasca. *Fonies Linguae Vasconum* 2.67-93. [Reprinted in Michelena 1987, 283-309].
- . 1972. A note on Old Labourdin accentuation. *Anuario del Seminario de Filología Vasca 'Julio de Urquijo'* 6.110-20. [Reprinted in Michelena 1987, 235-44].
- . 1977. *Fonética histórica vasca*, 2nd expanded edition. San Sebastián: Diputación de Guipúzcoa (Publicaciones del Seminario de Filología Vasca 'Julio de Urquijo').
- . 1987. *Palabras y textos*. Leioa: Univ. del País Vasco.
- TRASK, R. L. 1997. *The history of Basque*. London: Routledge.
- TXILLARDEGI [Alvarez Enparantza, José Luis]. 1984. *Euskal azentuaz*. Donostia: Elkar.

RESOLUTION ALGORITHMS FOR THE KOREAN REFLEXIVE *CAKICASIN*: A CATEGORIAL GRAMMAR APPROACH*

Yong-hun Lee

University of Illinois at Urbana-Champaign
ylee@uiuc.edu

The Korean reflexive *cakicasin* is known to be local, whereas another reflexive form, *caki*, allows long-distance antecedents. But, in some contexts, *cakicasin* has an antecedent that is located outside of the minimal S domain. That is, *cakicasin* is local, i.e., sentence-bound, in most cases, but permits a nonlocal antecedent sometimes. The goal of this paper is to provide more efficient computational algorithms for the Korean reflexive *cakicasin*, whether its antecedent is located within the minimal S domain or outside of the domain. For this purpose, Chierchia's (1988) Binding Theory is combined with Steedman's Combinatory Categorical Grammar (CCG). The combination will be called a CCG-like system. In this system, the [+refl] feature is instantiated as a feature of *cakicasin*, and it is percolated up until it meets a suitable antecedent. By these kinds of resolution algorithms, the syntactic behaviors of *cakicasin* are effectively implemented, reducing computational burdens considerably. In addition, we can easily contrast *cakicasin* with *caki*, saying that slightly different reflexive-antecedent pairs are responsible for the syntactic differences of these two reflexive forms. In sum, this paper develops more economical and unified algorithms that implement the Korean reflexive *cakicasin* using Categorical Grammar.

1. Introduction

Some languages have long-distance reflexives in addition to sentence-bound counterparts. Korean is one such language. As sentences from (1) to (4) illustrate, *caki* is a long-distance reflexive whereas *cakicasin* is sentence-bound.^{1,2}

* This is an extended version of Lee 2002b, focusing on *cakicasin* rather than *caki*. The system developed in this paper is actually implemented in JAVA as a partial fulfillment of Lee (forthcoming).

¹ In the examples below, relevant expressions are marked by boldface. Coreference relations between reflexive and its antecedent are represented by subscripted indexes. Superscripted numbers express preference rankings, making '1' the most preferred reading. These notational conventions come from Moon 1996.

² The preference rankings that are expressed in this paper are based on Moon's intuition. But, Kang's (1988) statistical approach supports this preference ranking. He counted occurrences of three reflexive forms, i.e., *caki*, *casin*, and *cakicasin*, in the KOREA-I corpus, which was built at Korea University. The data includes 5,000,000 tokens, where Case markers are attached to each word. The table in [7] illustrates the result. As this table shows, we have a total of 316 occurrences of *caki-lul*. Here, we have 165 cases of long-distance readings (52%), and 151 cases of sentence-bound readings (48%). Though there is little difference (4%), the long-distance reading is slightly preferred to the sentence-bound reading.

- (1) *Chelsoo*_i-ka *caki*_i-lul *salangha-n-ta*.
Chelsoo.NOM self.ACC love.PRES.DECL
'Chelsoo loves himself.'
- (2) *Chelsoo*_i-ka *cakicasin*_i-ul *salangha-n-ta*.
Chelsoo.NOM self.ACC love.PRES.DECL
'Chelsoo loves himself.'
- (3) *Chelsoo*_i¹-nun [*Younghee*_j²-ka *caki*_{ij}-lul *salangha-n-ta-ko*]
Chelsoo.TOP Younghee.NOM self.ACC love.PRES.DECL.COMP
sayngkakha-n-ta.
think.PRES.DECL
a. 'Chelsoo thinks that Younghee loves him.'
b. 'Chelsoo thinks that Younghee loves herself.'
- (4) *Chelsoo*_i-nun [*Younghee*_j-ka *cakicasin*_{ij}-ul *salangha-n-ta-ko*]
Chelsoo.TOP Younghee.NOM self.ACC love.PRES.DECL.COMP
sayngkakha-n-ta.
think.PRES.DECL
a. *'Chelsoo thinks that Younghee loves him.'
b. 'Chelsoo thinks that Younghee loves herself.'

In the paradigm of (1) and (2), *caki* and *cakicasin* make no difference, because both occupy the same positions and refer to the identical entity *Chelsoo*, which is located in the subject positions. But, in the paradigm of (3) and (4), both reflexives are clearly distinguished. In (3), *caki* can refer to either *Chelsoo* or *Younghee*, permitting both the long-distance reading and the sentence-bound one. In (4), *cakicasin* can refer to only the sentence-bound antecedent *Younghee*, not the long-distance antecedent *Chelsoo*.

But, in addition to these two paradigms, the following pair of sentences are also grammatically perfect in Korean.³

- (5) *Chelsoo*_i-nun [*caki*_i-ka *ttokttokha-ta-ko*] *sayngkakha-n-ta*.
Chelsoo.TOP self.NOM be smart.DECL.COMP think.PRES.DECL
'Chelsoo thinks that he is smart.'
- (6) *Chelsoo*_i-nun [*cakicasin*_i-i *ttokttokha-ta-ko*] *sayngkakha-n-ta*.
Chelsoo.TOP self.NOM be smart.DECL.COMP think.PRES.DECL
'Chelsoo thinks that he is smart.'

³ Jung-Min Jo (p.c.) suggested the following paradigm as a test for the subjecthood of *cakicasin*.

- i. *Elisek.keyto Chelsoo*_i-nun [*cakicasin*_i-i *ttokttokha-ta-ko*] *sayngkakha-n-ta*.
stupidly Chelsoo.TOP self.NOM be smart.DECL.COMP think.PRES.DECL
'Stupidly, Chelsoo thinks that he is smart.'
- ii. *Chelsoo*_i-nun *elisek.keyto* [*cakicasin*_i-i *ttokttokha-ta-ko*] *sayngkakha-n-ta*.
- iii. **Chelsoo*_i-nun [*cakicasin*_i-i *elisek.keyto* *ttokttokha-ta-ko*] *sayngkakha-n-ta*.
- iv. *Chelsoo*_i-nun [*cakicasin*_i-i *ttokttokha-ta-ko*] *elisek.keyto* *sayngkakha-n-ta*.
- v. *Chelsoo*_i-nun [*cakicasin*_i-i *ttokttokha-ta-ko*] *sayngkakha-n-ta*, *elisek.keyto*.

All the sentences except (iii) are acceptable. Sentence (iii) is impossible because the meaning of *elisekkeyto* is opposite to that of *ttokttokha-ta*. As the ungrammaticality of (iii) suggests, [*cakicasin*_i-i *ttokttokha-ta-ko*] becomes an island from the outside. *Cakicasin*-i is the subject of the subordinated clause, and its antecedent *Chelsoo* is located outside of this minimal S domain.

In (5) and (6), both reflexives occupy the subject positions of subordinated clauses, and they refer to *Chelsoo*, which is located outside of the minimal S domain. (5) is explained easily because *caki* allows a long-distance antecedent. But, the behavior of *cakicasin* in (6) is a little problematic in the sense that this reflexive usually takes a local sentence-bound antecedent. But, as Kang's (1998) corpus analysis in (7) also demonstrates, *cakicasin* sometimes allows antecedents that are located outside of the minimal S domain, though it is local in most cases.

(7) Distance of Antecedents for Korean Reflexives (Kang 1998:179)

	<i>Caki-lul</i>	<i>Casin-ul</i>	<i>Cakicasin-ul</i>
Local	151	311	66
Long-distance	165	123	5
Total	316	434	71

(*caki* - *casin* : $X^2 = 44.062$, $p < 0.001$)

In (7), *cakicasin* has 5 cases out of 71 examples (about 7%), where it has a nonlocal reading.⁴

As a result, we have three different distributions for *cakicasin*, i.e., (2), (4), and (6). In (2) and (4), its antecedent is local, while it is long-distance in (6). The goal of this paper is to provide more efficient computational algorithms for the Korean reflexive *cakicasin*. Here, the term *computational* has dual meaning. One is OPERATIONS ON REPRESENTATIONS à la O'Grady 1998 & 1999, and the other is COMPUTATIONAL IMPLEMENTATIONS. That is, the system introduced in this paper presupposes actual implementations on a computer. Accordingly, computational efficiency as well as theoretical discussion is crucial in this paper. For this purpose, this paper combines Chierchia's (1988) Binding Theory with Steedman's (1996 & 2000) Combinatory Categorical Grammar, which will be called a CCG-like system. In this system, *cakicasin* has [+refl] as a lexical feature, and this feature is percolated up until it meets a suitable antecedent. When it meets an antecedent, its feature is interpreted with that of the antecedent. By these kinds of resolution mechanisms, all the syntactic behaviors of *cakicasin* are effectively implemented, reducing computational cost considerably.

This paper is composed as follows. Section 2 discusses two possible types of implementations for *cakicasin*, and points out some problems that utilize Chomskyan traditions in implementations. Section 3 introduces basic ideas in Categorical Grammar, especially Steedman's (1996 & 2000) Combinatory Categorical Grammar (CCG), and Chierchia's Binding Theory. Section 4 develops a CCG-like system, and proposes resolution algorithms for *cakicasin* in this system. Section 5 discusses some advantages of resolution algorithms in the CCG-like system. Section 6 summarizes this paper.

⁴ The five cases here have to be understood as nonlocal, rather than really long-distance. The only instances where *cakicasin* can have an antecedent outside its own clause are when it is a subject of an embedded clause. Even then, it can only take an argument in the immediately superordinate clause as antecedent. So, unlike *caki*, *cakicasin* is not a true long-distance anaphor.

2. Some possible approaches in implementing *cakicasin*

In giving analyses of long-distant unbound reflexives in Korean, Japanese, Swedish, and Icelandic, Kang (1988) divided binding theories into two types. One is Domain-for-Antecedent Analysis (DA-account/DA-analysis), and the other type is Feature-Instantiation-and-Semantic-Interpretation Analysis (FI-account/FI-analysis). As its name implies, in the first type of analysis, reflexives are resolved within certain syntactic domains where both the reflexives and their antecedents are included. Chomskyan traditions such as GB (Chomsky 1981) belong to this type. In the second type of approach, reflexives are treated with feature instantiation and its semantic interpretation. That is, the reflexive feature is instantiated in each lexical entry, and this feature is percolated up until it meets a suitable antecedent. When a reflexive meets an antecedent, this reflexive feature is charged out with that of the antecedent. Analyses in GPSG and Categorical Grammar belong to this type.

If we computationally implement syntactic distributions of *cakicasin* by the DA-accounts, there may be roughly two types of approaches. The first one is to treat the sentence-bound reading as an anaphor, and the nonlocal reading as a pronominal. That is, *cakicasin* in (2) and (4) is an anaphor, but that of (6) is a pronominal. But, in this type of implementation, we would have to have two separate resolution algorithms for the same lexical item for *cakicasin*. Then, we need an additional algorithm that decides which resolution algorithm must be applied in the given situation. It increases time complexities in the implementations, which is an undesirable result.

The second type of possible approach is to define one maximal Binding Domain in which both readings are available. But, in order to define an appropriate Binding Domain for *cakicasin*, we need additional stipulations similar to Chomsky's (1981) accessible SUBJECT.

- (8) Governing Category (GC; Chomsky 1981:211)
 - a. AGR is co-indexed with the NP it governs.
 - b. β is a *governing category* for α if and only if β is the minimal category containing α , a governor of α , and a SUBJECT accessible to α .
- (9) Accessibility (Chomsky 1981:212)

α is *accessible to* β if and only if β is in the c-command domain of α and assignment to β of the index of α would not violate i-within-i condition.
- (10) i-within-i Condition (Chomsky 1981:212)

*[γ ... δ ...], where γ and δ bear the same index.

Along with these definitions, the following two sentences in English are explained as follows.

- (11) a. **John**_i thinks that [[a picture of **himself**_i] would be on sale].
- b. **John**_i thinks that [[**Bill**_j's picture of **himself**_{ij}] would be on sale].

In (11b), *himself* refers to *Bill*, not *John*, because the accessible SUBJECT of *himself* is *Bill*, and the GC that includes *himself* and its accessible SUBJECT is the subordinated clause, *Bill's picture of himself would be on sale*. In (11a), the subordinated clause cannot be a GC, because it violates the i-within-i condition in (10). Consequently, the GC for *himself* is extended to the matrix clause, and *himself* picks up *John* as its antecedent.

These mechanisms can be applied to *cakicasin* similarly, but with some modifications. Unlike English, Korean lacks AGR. Therefore, only the subject NP can be an accessible SUBJECT. When a reflexive cannot find an antecedent within the minimal S domain where the reflexive is included, the Binding Domain is extended into the immediate superordinate S domain and a GC is defined there. Thus, we can define the GCs in Korean without the i-within-i condition in (10). Now, let us take the three sentences, (2), (4), and (6), again, and see how these definitions work.

(2) *Chelsoo*_i-*ka cakicasin*_i-*ul salangha-n-ta*.
 Chelsoo.NOM self.ACC love.PRES.DECL
 'Chelsoo loves himself.'

(4) *Chelsoo*_i-*nun [Younghee*_j-*ka cakicasin*_i-*ul salangha-n-ta-ko]*
 Chelsoo.TOP Younghee.NOM self.ACC love.PRES.DECL.COMP
sayngkakha-n-ta.
 think.PRES.DECL
 a. *'Chelsoo thinks that Younghee loves him.'
 b. 'Chelsoo thinks that Younghee loves herself.'

(6) *Chelsoo*_i-*nun [cakicasin*_i-*i tokttokha-ta-ko]* *sayngkakha-n-ta*.
 Chelsoo.TOP self.NOM be smart.DECL.COMP think.PRES.DECL
 'Chelsoo thinks that he is smart.'

Example (2) is simple, since it has no subordinated clause. Only *Chelsoo* is accessible to *cakicasin*. Therefore, the S domain becomes the GC for *cakicasin*, in which it refers to *Chelsoo*. In (4), the subordinated clause, i.e., *Younghee-ka cakicasin-ul salangha-n-ta*, becomes the GC, where *cakicasin* can be resolved with *Younghee*. It cannot refer to *Chelsoo*, because the latter is outside of the GC for the former. In (6), the subordinated clause, i.e., *cakicasin-i tokttokha-ta*, cannot be a GC because it does not contain an accessible SUBJECT for *cakicasin*. Accordingly, the domain is extended up to the matrix clause, and the reflexive can find an antecedent, *Chelsoo*, in this domain.

So far, so good! All three sentences in (2), (4), and (6) are explained successfully. But, this type of approach also increases time and space complexity in the implementations considerably. If we implement resolution algorithms of *cakicasin* based on the GB framework, additional algorithms would be necessary for defining GC and implementing accessibility, in addition to the resolution algorithm itself. Since the GCs are configurationally defined in the tree structures, we need additional time and space for traveling those tree structures. Also, we need more processing time to pinpoint the position of the reflexive and its antecedent within the given GC. Consequently, implementations of resolution algorithms for *cakicasin*

within the GB framework require much time and space complexity, reducing computational efficiency. If we may have much simpler algorithms that are more efficient than those in this technique, it is reasonable to choose them, rather than more complicated ones.

Though we would tolerate those problems in the implementations, there is one big problem common to all of these approaches. The problem is that we have to spend too much in constructing and handling the tree structures. In DA-implementations, all the searching algorithms for an antecedent are performed on the tree structures. Accordingly, parsing algorithms such as Stabler's (1992) or Lee's (1996) algorithms are inevitably necessary. But, these parsing technologies require a great deal of memory space for storing each node in the tree structures and searching algorithms for traversing nodes in the tree structures. All of these things increase computational burdens, reducing the efficiency in the implementations. Besides, we have functional categories such as IP or CP and any other X'-level projections, and they also increase time and space complexity in constructing and traversing the tree structures.

Implementation of *cakicasin* in Chomsky's (1995) Minimalist Program (MP) has similar problems. Resolution algorithms in the MP may be similar to the FI-approaches in that lexical features play crucial roles in the system. But, those algorithms also presuppose a parsing technology such as that of Stabler 1999 & 2002. Accordingly, similar problems may be raised for the implementation. In addition, MP has more functional projections than GB, such as TP, NegP, Agr_sP, or Agr_oP. Handling these and their intermediate X'-level nodes increase time and space complexity significantly. As a result, we may spend much more time in constructing and traversing the tree structures rather than in resolution algorithms themselves.

As said above, we have a few different types of implementations for the reflexive *cakicasin*. Up to now, we have looked into possible DA-accounts, and pointed out some problems in their implementations. But, we did not discuss the FI-type accounts yet. The resolution algorithms developed in this paper are basically an FI-type approach. They will be discussed in detail in Section 4, after a brief introduction to Categorical Grammar and Chierchia's (1988) Binding Theory in Section 3.

3. Binding theory in Categorical Grammar⁵

Categorical Grammar was first introduced by Ajdukiewicz (1935) and later modified and advanced by Bar-Hillel, Curry, and Lambek. In this framework, we have two basic categories *n* and *s*, and the other categories come from the combinations of these two categories. In this system, all the syntactic phenomena are described and analyzed by functor-argument relations of constituents whose category information is stored in each lexical entry.

Steedman (1996 & 2000) extended previous studies in Categorical Grammar and developed Combinatorial Categorical Grammar (CCG). The most important

⁵ This section is also included in Lee 2001, 2002a, and 2002b.

characteristic of his system is that predicate-argument relations are projected by the combinatory rules of syntax, and that other operations are based on these relations (Steedman 2000:38). The most fundamental combinatory operation in Categorical Grammar is *functional application*, which is delineated in (12).

(12) Functional Application (Steedman 1996:13, 2000:37)

- a. $X/Y \quad :f \quad Y \quad :a \Rightarrow X \quad :fa \quad (>)$
 b. $Y \quad :a \quad X \setminus Y \quad :f \Rightarrow X \quad :fa \quad (<)$

Here, f refers to the semantic interpretation of the functor category and a to that of the argument category. As you can see, the semantic interpretation of the result is calculated by fa .

Chierchia used Categorical Grammar to explain Binding phenomena in English, and he described syntactic constraints of reflexives and pronominals as follows.

(13) Binding Theory in Categorical Grammar (Chierchia 1988:134)

- a. A reflexive must be bound to an F-commanding argument in its minimal NP or S domain.
 b. A non-reflexive pronoun must not be co-indexed with anything in its minimal NP or S domain.
 (F-command is simply c-command at function-argument structure.⁶)

Agreement in *number* and *gender* must hold between pronouns and antecedents, so pronouns can refer to their antecedents. The agreement checking algorithm is stated in (14a). $FT(n)$ in (14b) has three pieces of information: n is the index of the given NP, $gndr$ is gender, and $nmbr$ is number.

(14) Agreement between Antecedent and Pronouns (Chierchia 1988:132)

- a. $FT(n) \approx FT(m)$: The features associated with n are non-distinct from those associated with m .

b.
$$FT(n) = \begin{pmatrix} n \\ gndr \\ nmbr \end{pmatrix}$$

For example, the FTs of the three different NPs *John*, *himself*, and *her* can be stated as follows.⁷

⁶ This condition has different predictions for the sentences in (i) and (ii). It rules in (i), but rules out (ii) (Chierchia 1988:135).

- i. Mary showed the men each other.
 ii. *Mary showed each other the men.

Binding Theory before GB (Chomsky 1981) says that both of the sentences are grammatical, which is wrong. Tripartite structure is possible both for (i) and (ii), and *each other* and *the men* c-command each other in those structures. But, according to Larson's (1988) analyses with VP-shells, (i) and (ii) can be clearly distinguished from each other, because *the men* c-commands *each other* in (i), but the latter does not c-command the former in (ii). It is also pointed out in Lee 2001, 2002a, & 2002b.

⁷ Note that *John*, *himself*, and *her* stand for an R-expression, a reflexive, and a pronominal respectively.

(20) Six Attributes

- a. PHON
 - phonological/morphological form
 - concatenates a word to a stream of words
- b. CAT
 - has categorial information
 - such as S, NP, SNP, and so on
- c. AGR
 - agreement feature
 - index, type, gender, and number
- d. TRANS
 - semantic interpretation
 - based on Montagovian semantics
- e. NPS (NP Index Store)
 - something like a Cooper-storage
 - has indices of NPs including pronouns and their antecedents
- f. SLASH
 - similar to that of HPSG, except that it deals with pronouns
 - necessary to deal with crossover phenomena

Here, note that NPS (NP Index Store) substitutes for Chierchia's LPS (Local Pronoun Store). As you can see in the analyses in (17) and (18), the LPS is empty in the S nodes. Chierchia's LPS is LOCAL in the sense that the indexes in the LPS are available only within the minimal S domain. But, this raises problems in explaining Korean data. If [+refl] cannot be percolated up beyond the S node, *cakicasin* in sentence (6) and *caki* in sentence (3) cannot be explained. Therefore, we will use NPS instead of LPS in this paper, in the sense that the store is not local, i.e., not strictly limited into the minimal S domain.

The FUNCTIONAL APPLICATION on the CAT values triggers operations on TRANS and NPS values, and all the reflexives are resolved by these operations. In (19), AGR and SLASH are in parentheses, because the values for these two attributes will be omitted from the actual representations.

In the CCG-like system, the index types of NPs are classified as in (21), and this information is put into the FT specification of each NP.

(21) Three Index Types of NPs¹²

- a. $n_{[+name]}$: proper nouns
- b. $n_{[+refl]}$: reflexives
- c. $n_{[+pron]}$: pronominals

Following this change, the agreement checking algorithm in (14) has to be revised as in (22).

¹² In addition to those three types, we can add $n_{[+noun]}$ for COMMON NOUNS and $n_{[+mass]}$ for MASS NOUNS. But they are not included here for convenience.

(22) Agreement between Antecedent and Pronouns

a. $FT(n) \approx FT(m)$: The features associated with n are nondistinct from those associated with m .

b.

$$FT(n) = \begin{pmatrix} n \\ \text{type} \\ \text{gndr} \\ \text{nمبر} \end{pmatrix}$$

For example, the FTs of the three different NPs *Chelsoo*, *caki*, and *ku* are represented as follows.

(23)

$$\begin{array}{ccc}
 \text{Chelsoo}_1 & \text{caki}_2 & \text{ku}_3 \\
 FT(1) = \begin{pmatrix} 1 \\ +\text{name} \\ \text{male} \\ 3 \end{pmatrix} & FT(2) = \begin{pmatrix} 2 \\ +\text{refl} \\ [\quad] \\ 3 \end{pmatrix} & FT(3) = \begin{pmatrix} 3 \\ +\text{pron} \\ \text{male} \\ 3 \end{pmatrix}
 \end{array}$$

Note that the *gender* information of *caki* is underspecified. This underspecification reflects the fact that *caki* can refer to either a male entity, *Chelsoo*, or a female entity, *Younghee*, as in (3).

For the complementizer *-ko*, we will suppose that the **avop** of this lexical item is that of (24).

(24) **avop** for *-ko*
 $\langle ko, S \setminus S, \lambda\phi(\phi), NPS:\emptyset, SLASH:\emptyset \rangle$

The last step in the reflexive resolution algorithms is to revise Chierchia's algorithms in (16) so that Korean reflexives can be analysed properly. Example (25) shows the reflexive resolution algorithms after this revision. Step (16c) is eliminated, because we will not use this category combinatorics rule.

(25) Reflexive Resolution Algorithms in Korean

a. $NP_0 + (S \setminus NP_1) \setminus NP_2 \Rightarrow S \setminus NP_n$

- conditions: (i) $NPS(0) \cap NPS(1) = \emptyset$ noncoreference
 (ii) $SLASH(2) \cap (NPS(0) \cup NPS(2)) = \emptyset$ crossover
 (iii) $NPS(2) = NPS(0) \cup NPS(1)$ NPS-percolation
 (iv) $SLASH(2) = SLASH(0) \cup SLASH(1)$ slash-percolation

b. $NP_n + S \setminus NP_n \Rightarrow S$

- conditions: (i) $SLASH(2) \cap (NPS(1) \cup NPS(2)) = \emptyset$ crossover
 (ii) $NPS(2) = NPS(0) \cup NPS(1)$ NPS-percolation
 (iii) $SLASH(2) = SLASH(0) \cup SLASH(1)$ slash-percolation

c. Reflexives

(i) $A \Rightarrow A$
 $n [+refl] \in NPS \quad n \notin NPS$

- (ii) conditions: (a) $A = S \setminus NP$
 (b) $FT(A) \approx FT(n)$

(iii) translation: $\lambda x_n [A'(x_n)]$

There are three changes in (25) compared with the algorithms in (16). First, the head position is swapped. NP follows TV in (16), whereas it precedes ($S \setminus NP_n$)NP (=TV) in (25a). This is the result of head parameters. Note that English is a head-initial language, but that Korean is a head-final one. Second, we have the category ($S \setminus NP_n$)NP instead of TV. This is just for notational convenience. Third, the conditions in (16bi) and (16biv) are deleted in (25b). These deletions are necessary in order to permit the nonlocal readings.

4.2. An analysis of *cakicasin* in the CCG-like system

In order to account for the syntactic behaviors of *cakicasin* in the CCG-like system, it is necessary to view binding theories from a little different point of view. Instead of defining the Binding Domain (or GC) for *cakicasin* configurationally, we may say that property of this reflexive can be characterized as in (26).

(26) Characteristics of *cakicasin*

The reflexive *cakicasin* refers to the closest antecedent in the NPS store whose *person* and *number* information is compatible with that of *cakicasin*, whether its antecedent is located within the minimal S domain or outside of it.

That is, rather than saying that *cakicasin* must be bound within a certain syntactic domain, we define this reflexive as resolved with THE CLOSEST ANTECEDENT IN THE NPS STORE. If the index for a reflexive is k and we have two possible antecedent indexes i and j in the NPS store, NP_i is selected as an antecedent for the reflexive NP_k iff $|k-i| < |k-j|$. Since we do not need to define the syntactic domain for *cakicasin*, resolution algorithms are naturally FI-based. The feature [+refl] is INSTANTIATED as a lexical feature of *cakicasin*, and this feature is percolated up until it meets a suitable antecedent. When this feature meets an antecedent, it is SEMANTICALLY INTERPRETED with that of the antecedent. Resolution algorithms for Korean reflexives are enumerated in (25). In addition to them, the CCG-like system has one more operation, which is called the *reflexive-antecedent pairing algorithm*, in order to handle Korean reflexives properly. This algorithm makes a reflexive and its antecedent a pair. For example, for the two sentences (3) and (4), we may assume the reflexive-antecedent pairs in (27) and (28) respectively.

(3) *Chelsoo*₁-*nun* [*Younghee*₂-*ka cak*_{ij}-*lul salangha-n-ta-ko*]
 Chelsoo.TOP Younghee.NOM self.ACC love.PRES.DECL.COMP
sayngkakha-n-ta.
 think.PRES.DECL

- a. 'Chelsoo thinks that Younghee loves him.'
 b. 'Chelsoo thinks that Younghee loves herself.'

(4) *Chelsoo*_i-*nun* [*Younghee*_j-*ka cakicasin*_{ij}-*ul salangha-n-ta-ko*]
 Chelsoo.TOP Younghee.NOM self.ACC love.PRES.DECL.COMP
sayngkakha-n-ta.
 think.PRES.DECL

- a. * 'Chelsoo thinks that Younghee loves him.'
 b. 'Chelsoo thinks that Younghee loves herself.'

(27) Reflexive-Antecedent Pairs for Sentence (3)

a. *caki-lul* = *Chelsoo-nun*b. *caki-lul* = *Younghee-ka*

(28) Reflexive-Antecedent Pairs for Sentence (4)

a. *cakicasin-ul* = *Younghee-ka*

The analyses for these sentences proceed according to these reflexive-antecedent pairs. That is, the reflexive-antecedent pairs in (27) say that we have two possible cases for *caki* resolution. In the first case, i.e., (27a), *caki* is resolved with *Chelsoo-nun*, which results in the long-distance reading. In the second case, i.e., (27b), *caki* is resolved with *Younghee-ka*, which results in the sentence-bound reading. In (28), however, we have only one case for *cakicasin*, and this reflexive is resolved with only *Younghee-ka*, which is the sentence-bound reading.

Two naturally-raised questions are (i) how we state the reflexive-antecedent pairing algorithm, and (ii) how we can compute reflexive-antecedent pairs for each reflexive correctly, i.e., how we distinguish *cakicasin* from *caki*. For the first question, the intended reflexive-antecedent pairing algorithm can be stated as in (29).

(29) Reflexive-Antecedent Pairing Algorithm

For the reflexive *r* and all the potential antecedents *a*, make a pair '*r* = *a*' if $FT(r) \approx FT(a)$.

According to this algorithm, the reflexive-antecedent pairs for (3) and (4) are calculated as in (30) and (31).

(30) Reflexive-Antecedent Pairs for Sentence (3) (Tentative)

a. *caki-lul* = *Chelsoo-nun*b. *caki-lul* = *Younghee-ka*

(31) Reflexive-Antecedent Pairs for Sentence (4) (Tentative)

a. *cakicasin-ul* = *Chelsoo-nun*b. *cakicasin-ul* = *Younghee-ka*

Note that (27) and (30) are identical. That is, we get desirable results for the reflexive *caki*. But, (31) is different from (28). Example (31) has two cases under which *cakicasin* is resolved with its antecedent, whereas (28) has only one case. Then, how can we get the correct result, (28), from (31)? The answer is simple. The lexical characteristics of *cakicasin* in (26) play a role at this stage, and (28a) is deleted from the candidates, as in (32).

(26) Characteristics of *cakicasin*

The reflexive *cakicasin* refers to the closest antecedent in the NPS store whose *person* and *number* information is compatible with that of *cakicasin*, whether its antecedent is located within the minimal S domain or outside of it.

(32) Reflexive-Antecedent Pairs for Sentence (4) (Final)

a. ~~*cakicasin-ul* = *Chelsoo-nun*~~b. *cakicasin-ul* = *Younghee-ka*

Now, note that (32) is identical with (28). We get desirable results, (30) and (32), which are the same as (27) and (28) respectively. Since we say that the analyses proceed following these reflexive-antecedent pairs, we reach the correct results: *caki* in (3) can refer to either *Chelsoo* or *Younghee*, but *cakicasin* in (4) refers to only *Younghee*.

Now that we have all the necessary algorithms, let us take the three sentences (2), (4), and (6), and see how *cakicasin* is resolved in each case. First, sentence (2).

- (2) *Chelsoo*₇-*ka* *cakicasin*₇-*ul* *salangha-n-ta*.
 Chelsoo.NOM self.ACC love.PRES.DECL
 'Chelsoo loves himself.'

The only possible antecedent for *cakicasin* is *Chelsoo*, and *person* and *number* information of the former is compatible with that of the latter. Following the algorithm in (29), the reflexive-antecedent pair for this sentence is calculated as in (33).

- (33) Reflexive-Antecedent Pairs for Sentence (2)
cakicasin-ul = *Chelsoo-ka*

According to this reflexive-antecedent pair, sentence (2) is analyzed as in (34) overleaf.¹³

After *cakicasin-ul* and *salangha-n-ta* are combined, the Reflexive Rule in (25c) is applied to the $S\backslash NP_1$ node. By (25ci), the [+refl] index, 2_{+refl} , is erased from the NPS of this node. Because this category is $S\backslash NP$ and $FT(S\backslash NP) \approx FT(NP)$, the two conditions in (25cii) are satisfied. By (25ciii), the semantic interpretation of this node is changed from $love'(x_2)$ to $\lambda x_2[love'(x_2, x_2)]$, where A' is $love'(x_2)$.¹⁴ After *Chelsoo-ka* combines with *cakicasin-ul salangha-n-ta*, *cakicasin* is resolved with *Chelsoo*, resulting in $love'(c, c)$ from $\lambda x_2[love'(x_2, x_2)](c)$ by λ -conversion.

Next, let us go to the next sentence, sentence (4). Example (35) is an annotated form of (4), where bracketed superscripted numbers stand for phonological/morphological forms of each lexical item.

- (35) ^[1]*Chelsoo*₁-*nun* ^[2]*Younghee*₂-*ka* ^[3]*cakicasin*_{vj}-*ul*
 Chelsoo.TOP Younghee.NOM self.ACC
^[4]*salangha-n-ta-ko* ^[5]*sayngkakha-n-ta*.
 love.PRES.DECL.COMP think.PRES.DECL
 a. * 'Chelsoo thinks that Younghee loves him.'
 b. 'Chelsoo thinks that Younghee loves herself.'

¹³ Analysis of *caki* in sentence (1) is similar to (34), except that *caki-lul* substitutes for *cakicasin-ul* in (34). For detailed analysis see Lee 2002a.

¹⁴ Note that $\lambda x_2[love'(x_2, x_2)]$ is equal to $\lambda x_2[love'(x_2)(x_2)]$, by relational notation. For this convention, see Dowty, Wall, & Peters (1981:164).

- (34) <Chelsoo-ka cakicasin-ul salangha-n-ta, S, love'(c,c), NPS:1_{+name}>
 <Chelsoo-ka, NP_{1,c}, NPS:1_{+name}> <cakicasin-ul salangha-n-ta, SNP<sub>1,λx₂[love'(x₂,x₂)], NPS:1>
 <cakicasin-ul salangha-n-ta, SNP₁, love'(x₂), NPS:1, 2_{-refl}>
 <cakicasin-ul, NP_{2, x₂}, NPS:2_{+refl}> <salangha-n-ta, (SNP₁)\NP, love', NPS:1></sub>
- (36) a. * <[1]+[2]+[3]+[4]+[5]+[6], S, think'(c,^love'(y,c)), NPS:1_{+name}, 2_{+name}>
 <[1], NP_{1,c}, NPS:1_{+name}> <[2]+[3]+[4]+[5]+[6], SNP<sub>1,λx₃[think'(x₃,^love'(y,x₃))], NPS:1, 2_{+name}>
 <[2]+[3]+[4]+[5]+[6], SNP₁, think'(^love'(y,x₃)), NPS:1, 2_{+name}, 3_{+refl}>
 <[2]+[3]+[4]+[5], S', love'(y,x₃), NPS:2_{+name}, 3_{+refl}> <[6], (SNP₁)\S', think', NPS:1>
 <[2]+[3]+[4], S, love'(y,x₃), NPS:2_{+name}, 3_{+refl}> <[5], S'\S, λφ[φ], NPS:∅>
 <[2], NP_{2, y}, NPS:2_{+name}> <[3]+[4], SNP<sub>2, love'(x₃), NPS:2, 3_{+refl}>
 <[3], NP_{3, x₃}, NPS:3_{+refl}> <[4], (SNP₂)\NP, love', NPS:2></sub></sub>

- (36) b. $\langle [1]+[2]+[3]+[4]+[5]+[6], S, \text{think}'(c, \wedge \text{love}'(y, y)), NPS:1_{+name}, 2_{-name} \rangle$
 $\langle [1], NP_1, c, NPS:1_{+name} \rangle$ $\langle [2]+[3]+[4]+[5]+[6], \text{SNP}_1, \text{think}'(\wedge \text{love}'(y, y)), NPS:1, 2_{-name} \rangle$
 $\langle [2]+[3]+[4]+[5], S', \text{love}'(y, y), NPS:2_{-name} \rangle$ $\langle [6], (\text{SNP}_1)S', \text{think}', NPS:1 \rangle$
 $\langle [2]+[3]+[4], S, \text{love}'(y, y), NPS:2_{+name} \rangle$ $\langle [5], S'S, \lambda\phi[\phi], NPS:\emptyset \rangle$
 $\langle [2], NP_2, y, NPS:2_{-name} \rangle$ $\langle [3]+[4], \text{SNP}_2, \lambda x_3[\text{love}'(x_3, x_3)], NPS:2 \rangle$
 $\langle [3]+[4], \text{SNP}_2, \text{love}'(x_3), NPS:2, 3_{+ref} \rangle$
 $\langle [3], NP_3, x_3, NPS:3_{+ref} \rangle$ $\langle [4], (\text{SNP}_2)\text{NP}, \text{love}', NPS:2 \rangle$
- (39) $\langle [1]+[2]+[3]+[4]+[5], S, \text{think}'(c, \wedge \text{smart}'(c)) \rangle$, $NPS:1_{+name}$
 $\langle [1], NP_1, c, 1_{+name} \rangle$ $\langle [2]+[3]+[4]+[5], \text{SNP}_1, \lambda x_2[\text{think}'(x_2, \wedge \text{smart}'(x_2))] \rangle$, $NPS:1 \rangle$
 $\langle [2]+[3]+[4]+[5], \text{SNP}_1, \text{think}'(\wedge \text{smart}'(x_2)) \rangle$, $NPS:1, 2_{+ref} \rangle$
 $\langle [2]+[3]+[4], S', \text{smart}'(x_2), NPS:2_{+ref} \rangle$ $\langle [5], (\text{SNP}_1)S', \text{think}', NPS:1 \rangle$
 $\langle [2]+[3], S, \text{smart}'(x_2), NPS:2_{+ref} \rangle$ $\langle [4], S'S, \lambda\phi[\phi], NPS:\emptyset \rangle$
 $\langle [2], NP_2, x_2, NPS:2_{+ref} \rangle$ $\langle [3], \text{SNP}_2, \text{smart}', NPS:2 \rangle$

By the algorithm in (29), we have the reflexive-antecedent pairs in (31). But, owing to the lexical characteristics of *cakicasin* in (26), (31a) is erased from the set, and we get the final result in (32).

(31) Reflexive-Antecedent Pairs for Sentence (4) (Tentative)

- a. *cakicasin-ul* = *Chelsoo-nun*
 b. *cakicasin-ul* = *Younghee-ka*

(32) Reflexive-Antecedent Pairs for Sentence (4) (Final)

- a. ~~*cakicasin-ul* = *Chelsoo-nun*~~
 b. *cakicasin-ul* = *Younghee-ka*

Following this reflexive-antecedent pair, sentence (4) is analyzed as in (36) above.¹⁵

Note that the analysis (36a) is not allowed. The analyses in (36) DO exactly match with the reflexive-antecedent pairs in (32). The analysis in (36a) corresponds to the reflexive-antecedent pair in (32a), and that of (36b) to the pair in (32b). Since (36a) is prohibited, only (36b) is a possible analysis for the sentence (4).

Now, let us go to the final sentence, (6), which raises a problem. (37) is the annotated form of sentence (6).

- (37) ^[1]*Chelsoo*_i-*nun* [^[2]*cakicasin*_i-*i* [^[3]*ttoktokha-ta*-^[4]*ko*]
 Chelsoo.TOP self.NOM be smart.DECL.COMP
^[5]*sayngkakha-n-ta*.
 think.PRES.DECL

'Chelsoo thinks that he is smart.'

Because *Chelsoo* is the only possible antecedent for *cakicasin*, we get only one pair of reflexive-antecedent relations in (38), which is calculated by the reflexive-antecedent pairing algorithm in (29).

- (38) Reflexive-Antecedent Pairs for Sentence (6)
cakicasin-ul = *Chelsoo-nun*

According to this reflexive-antecedent pair, analysis for sentence (6) proceeds as in (39) above.¹⁶

Note that the analysis in (39) is similar to that of (34). The only difference is that the antecedent *Chelsoo* is located within the same minimal S domain in (34) whereas it is located outside of the minimal S domain in (39). Because reflexive resolution algorithms in the CCG-like system are insensitive to configurationally-defined syntactic domains, the [+refl] feature can be percolated up beyond the lower S node until it meets a suitable antecedent, *Chelsoo*.

¹⁵ Analyses of sentence (3) are similar to those given in (36a-b) with the following two differences. The first one is a lexical difference. That is, *cakicasin-ul* must be replaced with *caki-lul*. The second, and the most important, difference is that the analysis in (36a), which is prohibited for *cakicasin*, is perfect in the analyses for *caki*. Not only is (36a) allowed, but also it is preferred, as numeric rankings indicate. For detailed analyses of *caki*, see Lee 2002a. For comparison of *caki* and *cakicasin*, see Lee 2002b.

¹⁶ The analysis of *caki* in sentence (5) is similar to that given in (39), except that *caki-ka* substitutes for *cakicasin-i* in (39).

The analyses in (34), (36), and (39) demonstrate that syntactic distributions of *cakicasin* in the three sentences, (2), (4), and (6), are successfully explained with the reflexive resolution algorithms in the CCG-like system. Especially, it is desirable that *cakicasin* is implemented by the same resolution mechanisms in all three cases, by the Reflexive Rule and by Resolution by λ -conversion. This is possible because reflexive resolution algorithms in the CCG-like system take basically an FI-approach, i.e., feature instantiation and its semantic interpretation.

5. Advantage of the CCG-like system

Two criteria for evaluating computational algorithms are time complexity and space complexity. *Time complexity* refers to how much time it takes to run a program. The less time it takes, the better it is. *Space complexity* refers to how much space is necessary to run a program. The less space it takes up, the better program it is.

As discussed in Section 2, we have two types of Binding theories: DA-accounts and FI-accounts. The syntactic distributions of *cakicasin* in (2), (4), and (6) demonstrate that it takes a nonlocal antecedent sometimes, though it has a sentence-bound reading in most cases. As said in Section 2, this characteristic property of *cakicasin* raises some problems for traditional DA-type implementations, because constructing and traversing tree structures require much computational cost in the implementations.

The approach that this paper takes is basically an FI-approach. The most important advantage of this approach is that we can save much time and memory space in implementations, improving computational efficiency. As discussed in Section 2, if we would implement resolution algorithms of *cakicasin* based on the DA-approaches, we would have to define the Binding Domain configurationally. Accordingly, additional algorithms would be necessary for defining the domain in addition to the resolution algorithm itself. Since the domains are configurationally defined in the tree structures, we need additional time and space for constructing and traversing those tree structures. Consequently, implementations of resolution algorithms of *cakicasin* within the DA-approaches require much time and space complexity. Resolution algorithms developed in this paper are based on feature instantiation and its semantic interpretations, rather than syntactic configurations. Consequently, Binding Domains need not be defined configurationally. Accordingly, it is not necessary to construct and maintain all the tree structures in the memory space. Also, since we have no tree structure in the memory space, we do not need to traverse the tree structures. Therefore, we can save much processing time that is required for the DA-type implementations.

There are two more things that reduce time and space complexity. First, in the system developed in this paper, after the **avop** values of the mother node are calculated based on the values of the two component nodes, **avops** of the components are deleted from the memory space. When this result **avop** meets another **avop**, we have the same calculation procedures recursively. Consequently, only three **avop** spaces are enough for the processing a sentence, two for component

nodes and one for the result node. Along with this technique, we are able to be insensitive to the length of input sentences. That is, no matter how many input strings we have, only three **avop** spaces are necessary. The algorithms utilize these three **avop** spaces recursively. If we would implement the resolution algorithms with the DA-style approaches, the more input strings we have, the more memory space would be necessary. The reason is that, in those approaches, we need to store all the tree structures in the memory spaces. Secondly, because functional categories are not used in the CCG-like system, we do not need to waste much time and space that are required for handling those functional categories. Conclusively, the algorithms developed in this paper are faster and more efficient than the DA-based algorithms.

The second advantage is that we have unified reflexive resolution algorithms for *cakicasin*. As analyses in (36), (36), and (39) demonstrate, we can implement *cakicasin* with the same resolution mechanisms, i.e., by [+refl] instantiation, Reflexive Rule, and Resolution by λ -conversion. This result is desirable since we have unified resolution algorithms for one lexical item.

Another advantage is that we can easily capture similarities and differences of *caki* and *cakicasin* in the computational implementations. The two reflexives are similar in that they are resolved with the same series of resolution algorithms, i.e., the Reflexive Rule and Resolution by λ -conversion. Their differences are properly captured by different reflexive-antecedent pairs. These differences between *caki* and *cakicasin* come from lexical characteristics of *cakicasin*, which are mentioned in (26). Different analyses that come from different reflexive-antecedent pairs between *caki* and *cakicasin* are natural results, rather than stipulations.

6. Conclusion

In this paper, we have examined the syntactic distributions of the Korean reflexive *cakicasin*, and developed computational algorithms that resolve this reflexive. We observed that implementing syntactic distributions of this reflexive by DA-type accounts requires additional computational complexity in constructing and defining the syntactic domain configurationally.

The resolution algorithms that this paper uses take basically an FI-approach. The feature [+refl] is instantiated as a feature of *cakicasin*, and this feature is percolated up until it meets a suitable antecedent. When it meets an antecedent, this feature is charged out and semantically interpreted with that of the antecedent. Analyses in (34), (36), and (39) demonstrate that *cakicasin* is resolved with unified resolution algorithms, i.e., by the Reflexive Rule and Resolution by λ -conversion, whether the antecedent is within the minimal S domain or outside of it. We also found that *caki* and *cakicasin*, whose differences derived from reflexive-antecedent pairs, are resolved with the same mechanisms.

In sum, this paper provides unified resolution algorithms for the Korean reflexive *cakicasin*, whose syntactic distributions raised some problems in imple-

mentation. I hope this paper can give us an opportunity to understand characteristics of reflexives.

REFERENCES

- AJDUKIEWICZ, Kazimierz. 1935. Die Syntaktische Konnexität. *Studia Philisophica* 1:1-27.
- CHIERCHIA, Gennaro. 1988. Aspects of a Categorical Theory of Binding. In Oehrle, Bach, & Wheeler (eds.), 125-51.
- CHOMSKY, Noam. 1981. *Lectures on Government and Binding*, Dordrecht: Foris.
- . 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- DOWTY, David, Robert WALL, & Stanley PETERS. 1981. *Introduction to Montague Semantics*. Dordrecht: R. Reidel.
- HONG, Sungshim, & Yong-hun LEE. 1994. On the implementation of pronominal readings using Binding Condition B. *Proceedings of the Korean Society for Cognitive Science* 443-59.
- KANG, Beom-Mo. 1988. Unbounded reflexives. *Linguistics and Philosophy* 11:415-56.
- . 1998. Grammar and the use of language: Korean reflexives caki, casin, and cakicasin. *Korean Linguistics* 31:165-204.
- LARSON, Richard. 1988. On the double object construction. *Linguistic Inquiry* 19:335-91.
- LEE, Yong-hun. 1996. On an efficient parsing method using X-bar Theory. *Proceedings of the Korean Society of Cognitive Science* 33-51.
- . 2001. Implementation of pronoun readings in English: A Categorical Grammar approach. *Korean Journal of English Language and Linguistics* 1.4:609-27.
- . 2002a. Implementation of long-distance reflexives in Korean. *Proceedings of the 16th Pacific Asia Conference on Language, Information, and Computation* (PACLIC-16) 296-307.
- . 2002b. Bound vs. long-distance reflexives in Korean: A Categorical Grammar approach. Paper presented at the Conference of the International Circle of Korean Linguistics (ICKL).
- . [Forthcoming]. Anaphora resolution algorithms with category: English and Korean. University of Illinois at Urbana-Champaign, Ph.D. dissertation in Linguistics.
- LEE, Yong-hun, & Chongwon PARK. 2001. An analysis of Korean multiple accusative constructions and its computational implementation. Paper presented at the Annual Conference of the Linguistic Society of Korea. Kyung-Joo, Korea.
- LEVIN, Richard (ed.). 1992. *Formal Grammar: Theory and Implementation*. Oxford: Oxford University Press.
- MOON, Seung-Chul. 1996. *An Optimality Approach to Long Distance Anaphors*. Seoul: Thaeaksa.
- OEHELE, Richard, Emmon BACH, & Deirdre WHEELER (eds.). 1988. *Categorical Grammars and Natural Language Structures*. Dordrecht: R. Reidel.

- O'GRADY, William. 1998. Korean Case: A computational approach. *Proceedings of the 11th International Conference on Korean Linguistics*. Honolulu.
- . 1999. Computing Korean. Paper presented at the POSCO Conference. University of Texas at Austin.
- PARK, Chongwon. 2001. Two types of valence and two types of functor-argument relations. *Harvard Studies in Korean Linguistics IX*: 619-29.
- . [Forthcoming]. Semantic and combinational valences for Korean Case markers and COMPs: A bi-directional analysis and two types of functor-argument relations. To appear in *Japanese-Korean Linguistics XI*.
- STABLER, Edward. 1992. Implementing Government Binding Theory. In R. Levine (ed.), 243-75.
- . 1999. Parsing and generation for grammars with movement. University of California at Los Angeles, MS.
- . 2002. *Computational Minimalism: Acquiring & Parsing Languages with Movement*. Oxford: Blackwell.
- STEEDMAN, Mark. 1996. *Surface Structure and Interpretation*. Cambridge, MA: MIT Press.
- . 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.

**AN ALTERNATIVE ANALYSIS TO THE TEMPORAL ADVERBIAL
WHEN-CLAUSE***

Theeraporn Ratitamkul
University of Illinois at Urbana-Champaign
ratitamk@uiuc.edu

This paper aims to point out a defect in Enç (1987)'s analysis of the temporal *when*-clause in English and to present an alternative analysis that better captures the correct interpretation of the *when*-clause. To begin with, the paper first examines Enç's theory of anchoring conditions for tense as extended to analyze temporal adjunct clauses. It will be shown that although Enç's hypothesis seems to work in certain instances of *when*-clause, there exist cases in which Enç's predictions fail. The paper thus attempts to provide an alternative analysis to the temporal adjunct *when*-clause, which can solve problems arising from Enç's theory. Moreover, the proposed analysis is applicable to the interpretation of other temporal adverbial clauses, such as *before*-clauses and *after*-clauses as well. It should be noted that while constructing a new way to analyze temporal adjunct clauses, the paper assumes Enç's treatment of tenses as referential expressions that must be anchored as well as her insight that interpretations of tenses are derived from their syntactic properties.

1. Enç's analysis of the temporal adverbial *when*-clause and its shortcomings

First of all, instead of viewing tense as a sentential operator, Enç believes that tense, similarly to a nominal expression, is referential in nature. Owing to the dependent property of tense, she proposes the Anchoring Principle, stating that each tense must be anchored. Furthermore, Enç stipulates the following anchoring conditions.

- (1) a. Tense is anchored if it is bound in its governing category, or if its local Comp is anchored. Otherwise, it is unanchored.
- b. If Comp has a governing category, it is anchored if and only if it is bound within its governing category.
- c. If Comp does not have a governing category, it is anchored if and only if it denotes the speech time.

Enç's work is intended to tackle the interpretation of tense in an embedded clause. She also claims that her theory makes correct predictions with regard to

* I would like to thank Peter Lasersohn for extensive and helpful discussion on earlier drafts and James Yoon for insightful comments on the syntactic aspect of this paper.

tenses in temporal clausal adjuncts as well. As an illustration, Enç provides an analysis of the sentence below. (Enç 1987:655)

- (2) John visited his aunt when he was in London.

According to the theory of tense anchoring, the adjunct tense cannot be bound in its governing category by the matrix tense. This is because the latter is not high enough to be a possible antecedent; the matrix tense does not c-command the lower tense¹. Hence, the adjunct tense can only be anchored through its local Comp. Since the local Comp has no governing category, it denotes the speech time. This suggests that the adjunct tense is not related to the matrix tense by anchoring, and the interpretation of the tense in the adjunct clause is independent of that in the matrix. Enç further asserts that the adjunct denotes the interval provided by its tense, and that the tense shares its second index with the matrix tense, assuming that every temporal expression carries a pair of indices. It can be seen that the relation between the matrix tense and the adjunct tense is established by means of a referential second index. The denotation of double indexing concerning temporal adjuncts is as follows.

- (3) For $\alpha_{\langle i, j \rangle}$ and $\beta_{\langle k, l \rangle}$, if $j=l$ and $\beta_{\langle k, l \rangle}$ is interpreted before $\alpha_{\langle i, j \rangle}$, then $\llbracket \alpha \rrbracket \subseteq \llbracket \beta \rrbracket$.

Enç says that the order of the interpretation of temporal expressions is according to the syntax. As a result, the adjunct tense, which is, according to Enç, interpreted first, serves as the antecedent of the matrix tense in a sense. The interval denoted by the matrix tense is also contained in the interval denoted by the adjunct tense. Given sentence (2), the interval in which John is in London is interpreted before the interval in which he visits his aunt, and the time of visiting is included in the time of his being in London.

Nevertheless, Enç's analysis cannot be applied to sentences with *will* under the tense node (I) in the matrix clause and present tense in the temporal adjunct clause like the following.

- (4) I will have three grandchildren when I am 64.

Enç's analysis poses a major problem because the tense in the *when*-clause, which is considered the antecedent of tense in the matrix clause, is a present tense denoting the speech time. The predicted interpretation is inaccurate when the matrix tense depends on the adjunct tense for its denotation. The time of having three grandchildren, which intuitively comes after the speech time, has to be included in the time of being 64, which is in the present, meaning the same as the speech time. The outcome is then contradictory and incoherent.

Another problem emerging from Enç's proposal concerns inclusion, as Enç herself has noticed in her paper. *When*-clauses do not always involve inclusion of the time in the matrix. The sentence below reveals this fact.

- (5) Sue will go to bed when her father comes home.

¹ Enç assumes that the temporal adjunct clause is a sister of I'. However, Professor James Yoon has pointed out that it can be that clauses headed by *when* are generated under VP.

It can be observed that the time of Sue's going to bed is not included in the time of her father's coming home. The meaning expressed is rather that Sue's going to bed happens approximately at the same time as her father's coming home.

2. An alternative analysis

Following Enç (1987), I adopt the idea that tenses are referential expressions, and hence have to bear temporal indices; the interval denoted by tense according to its temporal index functions as a temporal argument of a verb. I also agree that the syntactic structure plays an important role in determining interpretations of tenses. However, in contrast to Enç, I assume that tense in the temporal adverbial *when*-clause is interpreted with respect to tense in the main clause.² That is, the matrix tense in a sense functions as the antecedent for the adjunct tense. This idea goes in line with Ogihara (1996)'s. He states that tense in a temporal adverbial clause is embedded in the scope of its matrix clause tense. Even though for Ogihara, his claim that the interpretation of tenses in temporal adjunct clauses is based on tenses in matrix clauses is made in favor of SOT (sequence of tense) phenomena, the Japanese data provided can be viewed as supportive of the new analysis to be explored.³ Following are Ogihara's examples from Japanese which only deal with the temporal adverbial *before*-clauses and *after*-clauses. (Ogihara 1996:181)

- (6) a. Taro-wa [Hanako-ni au mae-ni] denwa-o si-ta.
Taro-TOP Hanako-DAT meet-PRES before phone-ACC do-PAST
'Taro called Hanako before he saw her.'
- b. Taro-wa [Hanako-ni at-ta ato-de] denwa-o si-ta.
Taro-TOP Hanako-DAT meet-PAST after phone-ACC do-PAST
'Taro called Hanako after he saw her.'
- c. Taro-wa [Hanako-ni au mae-ni] denwa-o su-ru.
Taro-TOP Hanako-DAT meet-PRES before phone-ACC do-PRES
'Taro will call Hanako before he sees her.'
- d. Taro-wa [Hanako-ni at-ta ato-de] denwa-o su-ru.
Taro-TOP Hanako-DAT meet-PAST after phone-ACC do-PRES
'Taro will call Hanako after he sees her.'

The data show that tense in the *before*-clause is constantly present while the only tense allowed in the *after*-clause is past tense. It is unlikely that present tense in *before*-clauses and past tense in *after*-clauses invariably denote an interval in

² I do not think that this assertion can in any way apply to non-clausal temporal adverbs such as *yesterday*, *this morning* and the like. It seems to me that intuitively, these adverbs serve as indicators for locating intervals denoted by tenses at some specific time. The matter of how the interpretation of non-clausal temporal adverbs can be achieved is not in the scope of this paper.

³ Like Ogihara, I think that the interpretation of tenses in temporal adverbial clauses is carried out in the same manner for both English and Japanese. Thus, instead of treating an adjunct tense in Japanese as depending on its matrix tense, while assuming that an adjunct tense in English is interpreted independently, I suggest that for both English and Japanese an adjunct tense rely on its matrix tense in order to be correctly interpreted. Professor Yoon has, nevertheless, suggested that the interpretations of tense in Japanese are derived from *ato* and *mae*, which are nominal in nature and are deictic, not from the verb morphology.

the present time and an interval in the past, respectively. The data suggest that the denotation of tense in a temporal adverbial clause relies on the interpretation of its matrix tense in some way.

In order to account for the semantics of the temporal adverbial *when*-clause, I adopt Stump (1985:127)'s proposal in which the English subordinating conjunctions *before*, *after*, *when*, and *while* are classified as basic expressions of the category TSC.

(7)	<u>Category</u> TSC	<u>Description</u> the basic category of temporal subordinating conjunctions	<u>Basic expressions</u> <i>before, after, when, while</i>
-----	------------------------	---	---

Assuming the analysis of tense as a temporal predicate, Stump stipulates that temporal adverbial clauses headed by conjunctions in the TSC category have either the temporal predicate NONPAST or PAST, with a condition that NONPAST is exclusive to temporal adverbial clauses only; it cannot occur in main clauses. I, however, will not follow Stump's analysis of temporal predicates to explain temporal adjunct clauses. It is puzzling to a certain degree to have the temporal predicate NONPAST simply apply to tenses in temporal adverbial clauses. In addition, as Stump points out, the rules he proposes allow a number of ungrammatical sentences like the following to be generated and more explanations are needed in order to rule out these types of sentences. (Stump 1985:144)

- (8) a. *John will leave when Mary arrived.
 b. *John left after Mary is singing.
 c. *John will leave after Mary arrived.
 d. *John left before Mary arrives.
 e. *John will leave when Mary will arrive.

On the contrary, I will continue to assume Enç's treatment of tenses as temporal indices and her view of a tense as denoting an interval that relates to an interval its local Comp denotes. According to Enç (1987)'s, the definitions of past tense and present tense are as follows.

- (9) a. Where α is a past tense, β is a Comp with a temporal index, and β is the local Comp of α , $\llbracket \alpha \rrbracket$ is an interval T such that every moment t in T precedes every moment t' in $\llbracket \beta \rrbracket$.
 b. Where α is a present tense, β is a Comp with a temporal index, and β is the local Comp of α , $\llbracket \alpha \rrbracket$ is an interval T such that $T = \llbracket \beta \rrbracket$.

Since Enç's theory does not address the interpretation of the modal *will* followed by a base verb, the definition of such sequence must be stated in order to tackle the problems occurring from Enç's analysis of the *when*-clause. It has been proposed that English has two tenses, namely past and present (e.g., Ogihara 1996, Smith 1978). Hence, treating *will* as denoting future tense seems to be out of place. In fact, Smith (1978:49) gives strong arguments against treating *will* as a tense.

- (10) a. *Will* can occur with Present and Past as well as Future sentences.
 (i) The store *will* have your book by now.
 (ii) The documents *will* have arrived last week.
 b. *Will* is not the only predictive form that appears in Future. There is no reason to give it a status different from *may*, for instance.
 c. *Will* and other modals occur with present and past tense.

Following that idea, the sequence *will* and a base verb can be considered present tense. However, it is obvious that an interval denoted by *will* + infinitive takes place after an interval denoting the speech time, whereas the simple present tense and the speech time share the same interval. In order to capture the discrepancies between the present tense with *will* and the simple present tense in this framework, it can be stipulated that the modal *will* forces an interval denoted by present tense to occur after that denoted by its local Comp. The definition of *will* is stated below.

- (11) The definition of *will*
 a. *Will* is a present tense with a temporal index PRESENT.
 b. Where α is a modal *will*, β is a Comp with a temporal index, and β is a Comp of α , $\llbracket \alpha \rrbracket$ is an interval T such that every moment t in T follows every moment t' in $\llbracket \beta \rrbracket$.

Moreover, indexation will be used to indicate that two intervals are simultaneous or of different time. Hence, tenses with temporal indices such as PRESENT or PAST will be coindexed if the intervals they denote occur approximately at the same time; they will carry the same additional index (i, j, or k). Introducing a new index helps to pinpoint the occurrences of two intervals, to see whether they are simultaneous or not, while having an index PRESENT or PAST is simply to locate an interval a tense denote as compared to an interval denoted by its local Comp. With an additional index, it is clear whether two intervals or part of them belong to the same interval. That is, two intervals share the same index if they overlap.

- (12) For α_i and β_j such that α and β are tenses or Comps, and $i=j$, there is an interval $\llbracket \gamma \rrbracket$ such that $\llbracket \gamma \rrbracket \subseteq \llbracket \alpha \rrbracket$ and $\llbracket \gamma \rrbracket \subseteq \llbracket \beta \rrbracket$.

The translation of *when* under this analysis is $\lambda P \lambda Q \exists t_i \exists t_j [P(t_i) \wedge Q(t_j) \wedge i=j]$ where t_i and t_j are intervals that overlap, and P and Q are verbs. It can be seen that *when* requires two intervals denoted by tenses and serving as arguments of verbs to be simultaneous. Next, in order to capture the dependency of the temporal adjunct tense on the matrix clause tense, I set forth the following stipulations.

- (13) Condition for temporal adverbial clauses
 TSC obtains its temporal index, PAST or PRESENT, from the immediately m-commanding I.

It follows from these conditions that TSC, which is in the local Comp position of an adjunct clause, will share the same index with its immediately m-commanding I. Furthermore, a constraint on temporal adverbial clauses must also be stated.

(14) Constraint on temporal adverbial clauses

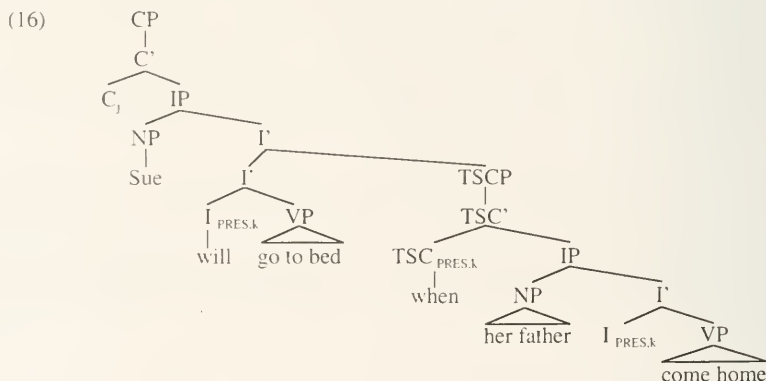
A temporal index of an adjunct tense must be the same as that of its local Comp, that is, TSC.

Consequently, tense is anchored through its local Comp in this way, in accordance with the Anchoring Principle proposed by Enç saying that tense must be anchored. Nonetheless, an extra condition must be added to the Anchoring Conditions.

(15) Additional Anchoring Condition

If Comp does not have a governing category, it is anchored if and only if it shares the same temporal index with its m-commanding I.

As an illustration, the analysis proposed here can apply to sentence (5), *Sue will go to bed when her father comes home*. The I of the matrix clause dominates *will*, and it has the temporal index PRESENT and an additional index k. Since the modal *will* constrains the interval denoted by the present tense to follow the interval the local Comp denotes, the two intervals do not take place at the same time. They thus have different indices. While the present tense with *will* has the index PRESENT,k its local Comp, which denotes the speech time, has a different index j (where $j \neq k$), showing that the two intervals do not overlap. The interval the present tense with *will* denotes takes place after the speech time. Next, the TSC *when* receives its PRESENT,k tense index from I, which immediately m-commands it; the first maximal projection of I with the PRESENT,k index dominates the TSC *when*. The adjunct tense in I will also have the PRESENT,k index, following the constraint proposed for temporal adjunct clauses. Hence, it can be seen that tenses in both the matrix clause and the adjunct clause denote overlapping intervals that come after the speech time, meaning the two eventualities simultaneously take place. This interpretation seems to go along well with the intuition. Presented below is the tree diagram for sentence (5) *Sue will go to bed when her father comes home*. Following Enç, a temporal adjunct headed by *when* is generated as a sister of I'. This makes sense because a temporal adjunct clause modifies a time in which another event takes place, and, as an adjunct, it is a daughter of an I'-projection.



Regarding the morphological restriction which prevents the occurrence of the modal *will* in adjunct clauses, Hornstein (1977), in light of Heinamaki (1974), stipulates that temporal conjunctions do not allow epistemic modals in the second clause. The following sentence is thus unacceptable.

(17) * I will have three grandchildren when I will be 64.

Taking the proposed analysis into consideration, the problem posed by Enç's is then solved. It is also noticeable that the present analysis not only eliminates the need for matrix tense and adjunct tense to share a second index, it also removes the problem of inclusion mentioned before. Instead, matrix clause tenses and adjunct clause tenses are related through their syntactic relation, in particular through relation of tenses and TSCs. The semantics of the *when*-clause can be thus described.

Following is the truth condition of a sentence containing the temporal adverbial *when*-clause.

(18) Truth condition for *when*-clauses

Where ϕ and ψ are sentences, ϕ *when* ψ is true at an interval i if and only if

- 1) ϕ is true at an interval j denoted by its tense, and
- 2) ψ is true at an interval k denoted by its tense, and
- 3) the interval i is included in both interval j and interval k .

For instance, for sentence (5) to be true, both the incident of Sue's going to bed and her father's coming home must be simultaneously true at an interval. This captures the intuition that eventualities in the two clauses, the matrix and the *when*-clause, take place at the same time.

3. An account of English morphology in temporal adverbial *when*-clauses

In light of the analysis proposed, the morphological facts of English displayed in temporal adverbial clauses become understandable. Since in English, the temporal index of an adverbial clause is the same as its local Comp and the matrix clause tense, it can follow that surface tenses in both clauses necessarily go hand in hand. For example, if the matrix tense is in present, the adjunct tense must also be in present. Likewise, when the matrix clause has past tense, the adjunct tense must be in the past as well. As mentioned before, the discrepancy between the matrix clause and the adjunct clause regarding the presence of *will* is due to a language-specific fact of English. Moreover, the choice of the present tense morpheme in the *when*-clause in such cases follows from the fact that the lower I denotes the present tense as it receives the PRESENT index from the matrix I. The acceptability versus the unacceptability of the following sentences reveals that the analysis of adjunct tenses as depending on matrix tenses seems to yield accurate predictions.

(19) a. Pat opens the umbrella when it rains.⁴

⁴The analysis here does not concern the habitual reading of the present tense.

- b. Pat opened the umbrella when it rained.
- c. Pat will open the umbrella when it rains.
- d. *Pat opens the umbrella when it rained.
- e. *Pat opened the umbrella when it rains.
- f. *Pat will open the umbrella when it rained.

Sentences (19d) to (19f) are ungrammatical because of morphological clashes. To begin with, as mentioned above, the temporal index a tense bears plays a role in determining the denotation of that tense, and the constraint on temporal adverbial clauses requires that the temporal index of tense in an adjunct clause be identical to its local Comp. Therefore, it can be derived that in (19d), the tense in the *when*-clause receives the index PRESENT from its local Comp, the TSC *when*, which gets the temporal index from the m-commanding I in the upper clause. The tense in the *when*-clause is, therefore, interpreted as denoting an interval in the present time. Nonetheless, it carries the past tense morpheme *-ed*, yielding disagreement between the interpretation and tense morphology. The sentence is thus unacceptable.

Likewise, (19e) shows another morphological conflict. Tense in the adverbial adjunct is interpreted as designating a past interval, and is supposed to have a past tense morpheme. However, the sentence is uninterpretable, since the adjunct tense appears with the present tense morpheme. As for (19f), the same reasoning applies. The sentence is ungrammatical due to the fact that the past tense morpheme appears where the present time interpretation is sought. Again, a morphological disagreement takes place.

It can be concluded that acceptable English sentences with temporal adverbial clauses like those in (19a) to (19c) reveal an isomorphic relation between matrix tenses and adjunct tenses, and also an agreement with regard to tense interpretation and tense morpheme. As having been shown, the proposed analysis gives correct predictions for acceptable and unacceptable sentences in English as well as successfully accounting for phenomena involving English tense morphology in temporal adverbial clauses.

4. Application to *before*-clauses and *after*-clauses

The analysis of the temporal adverbial *when*-clause presented in this paper can be extended to capture the semantics of other temporal adverbial clauses, including *before*-clauses and *after*-clauses. First of all, *before* can be translated as $\lambda P \lambda Q \exists t_i \exists t_j [P(t_i) \wedge Q(t_j) \wedge t_i < t_j]$, whereas *after* translates as $\lambda P \lambda Q \exists t_i \exists t_j [P(t_i) \wedge Q(t_j) \wedge t_j < t_i]$, where P and Q are verbs, and t_i is an interval functioning as an argument of the first verb, whereas t_j is an interval functioning as an argument of the second verb. Furthermore, as previously assumed, tense, being a referential expression, must be anchored, according to Enç's Anchoring Principle. As suggested under the proposed analysis, tense in a temporal adjunct clause is dependent on tense in a matrix clause for its interpretation. How this dependency is carried out has already been presented. The outcome is that the matrix tense and the adjunct tense share

their temporal index. This, I assume, is also applied to *before*-clauses and *after*-clauses as to be illustrated henceforth.

The sentences in (20) can be used to support the idea that tenses in adverbial *before*-clauses and *after*-clauses have to be related to those in main clauses in order to be interpretable. Apart from forbidding the use of modal auxiliary in temporal adjunct clauses, thus not allowing the occurrence of *will*, the English language also has a constraint that tense morphemes must correspond to the interpretations of tenses. Otherwise morphological clashes will arise, making sentences unacceptable as seen in examples in (21).

- (20) a. John fed the cat before Mary arrived.
 b. John fed the cat after Mary arrived.
 c. John feeds the cat before Mary arrives.
 d. John feeds the cat after Mary arrives.
 e. John will feed the cat before Mary arrives.
 f. John will feed the cat after Mary arrives.
- (21) a. *John fed the cat before Mary arrives.
 b. *John fed the cat after Mary arrives.
 c. *John feeds the cat before Mary arrived.
 d. *John feeds the cat after Mary arrived.
 e. *John will feed the cat before Mary arrived.
 f. *John will feed the cat after Mary arrived.

It is worth noticing that in Japanese, there exists no restriction that forces tense morphemes to match with tense denotations. As seen in example (6) above, tense in the *before*-clause is unchangeably present while that in the *after*-clause is consistently past. This characteristic of tenses in *before*-clauses and *after*-clauses in Japanese can be attributed to the lexical properties of *before* and *after* that tolerate no other tense than present and past, respectively. If a tense morpheme in a temporal adverbial clause were to alter to be in accordance with the tense interpretation, the sentence would be ungrammatical. The examples below are from Ogihara (1996: 181-182).

- (22) a. *Taroo-wa [Hanako-ni at-ta mae-ni] denwa-o si-ta.
 Taro-TOP Hanako-DAT meet-PAST before phone-ACC do-PAST
 [Intended] 'Taro called Hanako before he saw her.'
- b. Taroo-wa [Hanako-ni au mae-ni] denwa-o si-ta.
 Taro-TOP Hanako-DAT meet-PRES before phone-ACC do-PAST
 'Taro called Hanako before he saw her.'
- c. *Taroo-wa [Hanako-ni au ato-de] denwa-o su-ru.
 Taro-TOP Hanako-DAT meet-PRES after phone-ACC do-PRES
 [Intended] 'Taro will call Hanako after he sees her.'
- d. Taroo-wa [Hanako-ni at-ta ato-de] denwa-o su-ru.
 Taro-TOP Hanako-DAT meet-PAST after phone-ACC do-PRES
 'Taro will call Hanako after he sees her.'

Truth conditions of *before*-clauses, and *after*-clauses are as follows.

(23) Truth conditions for *before*-clauses and *after*-clauses

Where ϕ and ψ are sentences:

- a. ϕ *before* ψ is true at an interval k if and only if
 - 1) ϕ is true at an interval i , and
 - 2) ψ is true at an interval j , and
 - 3) the interval i precedes the interval j , and
 - 4) both the interval i and the interval j are included in the interval k
- b. ϕ *after* ψ is true at an interval k if and only if
 - 1) ϕ is true at an interval i , and
 - 2) ψ is true at an interval j , and
 - 3) the interval j precedes the interval i , and
 - 4) both the interval i and the interval j are included in the interval k .

As an illustration, sentence (20e) is true at an interval that includes an interval in which John feeds the cat is true preceding an interval in which Mary arrives is true. The two intervals, as their temporal index PRESENT (with *will*) suggests, follow the interval denoted by the speech time in Comp. On the other hand, sentence (20f) is true if and only if the interval in which John feeds the cat is preceded by the interval indicating that Mary arrives, and both are included in a larger interval. Both intervals again come after the interval of speech time. Besides, tenses in *before*-clauses and *after*-clauses do not share the same additional index as tenses in matrix clauses. This is because the intervals they denote do not occur simultaneously.

It can be seen that the analysis proposed for *when*-clauses can be applied to *before*-clauses as well as *after*-clauses. Thus, the semantics of temporal adverbial clauses can be accounted for in a unified manner. In addition, the analysis is able to achieve interpretations of temporal adjunct clauses cross-linguistically as demonstrated in examples from English and Japanese.

5. Conclusion

It is clear that Enç's analysis poses problems in interpreting sentences containing the temporal adverbial *when*-clause. An alternative analysis is then proposed, maintaining her concept of tenses as temporal referential expressions, and her consideration of syntactic properties as generating interpretations for tenses. The crucial difference lies in the fact that the new analysis views matrix tense as the antecedent for temporal indexation, instead of adjunct tense. This assumption together with some modifications to Enç's statements plus certain independently motivated stipulations give a plausible explanation and correct predictions concerning interpretation of tenses not only in the temporal adjunct clauses headed by *when*, but also in other temporal adjuncts such as *before*-clauses and *after*-clauses. The analysis presented in this paper, hence, should be taken into consideration to account for the semantics of temporal adverbial clauses.

REFERENCES

- ENÇ, Murvet. 1987. Anchoring conditions for tense. *Linguistic Inquiry* 18:4.633-57.
- HEINÄMÄKI, Orvokki. 1974. Semantics of English temporal connectives. University of Texas at Austin, Ph.D. dissertation in Linguistics.
- HORNSTEIN, Norbert. 1977. Towards a theory of tense. *Linguistic Inquiry* 8:3.521-57.
- OGIHARA, Toshiyuki. 1996. *Tense, Attitudes, and Scope*. Dordrecht: Kluwer.
- SMITH, Carlota. 1978. The syntax and interpretation of temporal expressions in English. *Linguistics and Philosophy* 2.43-99.
- STUMP, Gregory. 1985. *The Semantic Variability of Absolute Constructions*. Dordrecht: Kluwer.

REVIEW

Dixon, R. M. W., and Alexandra Y. Aikhenvald: *Word: A Cross-Linguistic Typology*. Cambridge: Cambridge University Press, 2002. Pp. xiii +290. Price: (hardback) \$60.00, ISBN 0-521-81899-0.

José Ignacio Hualde
University of Illinois at Urbana-Champaign
j-hualde@uiuc.edu

This book consists of an introductory chapter by the editors, Dixon & Aikhenvald (henceforth D&A), nine chapters on the notion of word and related matters on individual languages and a final 'conclusions' chapter by P. H. Matthews. The nine chapters between the introduction and the conclusions are by the following authors and on the following topics: A. Aikhenvald on clitics, especially in Tariana, an Arawak language from Amazonia; A. Woodbury on the word in Cup'ik, a variety of Central Alaskan Yup'ik; J. Henderson on the word in Eastern/Central Arrernte, a language from central Australia; R. M. W. Dixon on the word in Jarawara, an Amazonian language; U. Zeshan on the notion of word in sign languages, with an emphasis on Indo-Pakistani Sign Language; R. Rankin et al. on the word in Siouan, from both a synchronic and a diachronic perspective; K. Olawsky on the word in Dagbani, a Gur language of northern Ghana; A. Harris on the word in Georgian; and B. Joseph on the word in Modern Greek.

D&A propose that the notion of phonological word needs to be conceptually distinguished from that of grammatical word and list a number of criteria that can be used for the identification of each of the two units. Depending on the language, grammatical and phonological words may coincide to different degrees, but, in their view, there is no a priori reason why morphological and phonological words should be coextensive in all languages all the time. A draft of Chapter 1 was sent to all contributors and most of them adopt this approach. Thus, most chapters contain sections on phonological words, morphological words, clitics ('something that is a grammatical word but not a complete phonological word'), and cases of mismatches between grammatical and phonological words (25). The exception to this theoretical stand (and chapter organization) is Joseph's chapter on the word in Greek. Although other authors have analyzed Greek as possessing a number of clitics, Joseph — following Zwicky (1985) and other work by this author — argues that the notions of word and affix are sufficient for the analysis of this and other languages, provided that we recognize that there are typical and atypical words and affixes. The term 'clitic' is used by other authors to refer to atypical cases (atypical words or atypical affixes) but, in Joseph's view, as in Zwicky's, nothing is gained by introducing a third category, clitic, as another tool in our box of analytical tools. In Joseph's chapter, unlike in most other chapters in the book, phonological and grammatical facts relevant for the definition of 'word' are dis-

cussed together. Personally, I find Joseph's view more appealing than that espoused in the other chapters. Whereas D&A are undoubtedly correct in stating that phonological and grammatical units do not necessarily coincide in all cases (and this is abundantly demonstrated in this book), I am not convinced that all evidence of one of these two types, phonological or grammatical, will necessarily converge in identifying the same word units in a given language. Let us consider one example. In the Basque dialect of Ondarroa, intervocalic /d/ has become a flap (this sound is represented as *r*, since the process has caused neutralization with the pre-existing rhotic flap). The 'flapping of /d/' is generally limited to word domains, but it also affects /d/-initial inflected verbs; compare *Fidel-da* 'it is Fidel', *Fidel-dator* 'Fidel is coming' with *Kól-do-ra* 'it is Koldo', *Kól-do-rator* 'Koldo is coming'. Since, as in all other Basque dialects and in Spanish (in which all Ondarroa Basque speakers are bilingual), the flap is otherwise excluded from the word-initial position, it seems that we must conclude that these inflected verb forms phonologically attach to the noun to their left, creating a single phonological word that contains two grammatical words. Converging evidence of 'phonological wordhood' is also provided, for instance, by a vowel assimilation process that raises word-final low vowels to mid after a high vowel, as in the example *Péru-re* 'it is Peru'. So far, so good: this would be yet another case of mismatch between phonological and morphological words. The problem is that not all the phonological evidence converges in identifying these verbs, *da* 'is', *dator* 'is coming', etc., as forming a phonological word unit with the noun to their left. In particular, the accentual evidence is in clear conflict with this. In Ondarroa Basque stems (and affixes) can be lexically accented or unaccented. If the stem is accented, an accent surfaces on the penultimate syllable of the word. For instance, the name *Kól-do* is accented. Its accent can be seen to shift to the penultimate syllable in inflected forms such as *Koldóna* 'the one of Koldo's', *Koldontzáko* 'for Koldo'. Crucially, however, following inflected verbs are outside of the accentual word domain, as shown by the examples *Kól-do-ra* 'it is Koldo', *Koldontzáko-ra* 'it is for Koldo', with word-level penultimate accent. The domains that are identified as phonological words in Ondarroa Basque will thus depend on the specific phenomena that are chosen for this classificatory purpose, not on a simpler distinction between phonological and grammatical 'word' domains. Typically all word-level phonological evidence will point in the same direction, just like typically phonological and grammatical evidence for the word will converge, but this is not a necessary state of affairs. I am thus more comfortable with the gradualistic view presented by Joseph in his chapter. It does not seem obvious to me that the replacement of a single notion of word by two, phonological word and grammatical word, as D&A propose, represents a clear methodological or conceptual improvement.

This book offers a wealth of information on the status of the 'word' in a typologically and geographically diverse set of languages, both oral and signed, including some with very striking mismatches between the word-size units which are defined according to different morphosyntactic and phonological criteria. For quite a few of the languages in this book, this sort of information cannot be found anywhere else. D&A have done an excellent job at giving coherence to the volume

by providing a preliminary version of the first chapter to all contributors, which they were able to use as a model for the description (to the extent that they did not disagree with D&A's basic premises). The conclusion chapter, by Matthews, which contains reference to many issues raised by the other contributors in their chapters (as well as providing some interesting reflections on how our grammatical tradition — including the concept of word — has been influenced by the specific morphological and syntactical properties of the Latin language) is also very useful and contributes to making this book much more than a collection of descriptions of the 'word' facts of a random set of languages.

The book has been carefully edited. I only noticed a single lapsus, in a footnote on p. 7: 'The *up* can move to the left over an object that is a full NP but not over a preposition — *I made up the story* but not **I made up it.*' It should be 'over a pronoun'.

REFERENCE

ZWICKY, Arnold. 1985. Clitics and particles. *Language* 61.283-305.

REVIEW*

Ernst, Thomas: *The Syntax of Adjuncts*. Cambridge: Cambridge University Press, 2002. Pp. vii + 555. Price: (hardback) \$90.00, ISBN 0-521-77134-X.

Cedric Boeckx
Harvard University
cboeckx@fas.harvard.edu

The syntax of adverbial adjuncts has played an important role in the development of syntactic theory. This is especially true in the Principles and Parameters model laid out in Chomsky 1981. For example, the ban on extraction out of adjuncts, as opposed to complements, was a key aspect of Huang's (1982) Condition on Extraction Domain, which has influenced much of the research on locality since then.

- (1) a. *who did you arrive [after Bill met *t*]
b. who did you see [pictures of *t*]

The fact that adjuncts are much harder to extract out of (weak) islands such as the *wh*-island (see Lasnik & Saito 1984 among many others) also played a significant role in defining conditions on syntactic relations.

- (2) a. *how did John wonder [what to fix *t*]
b. ?what did John wonder [how to fix *t*]

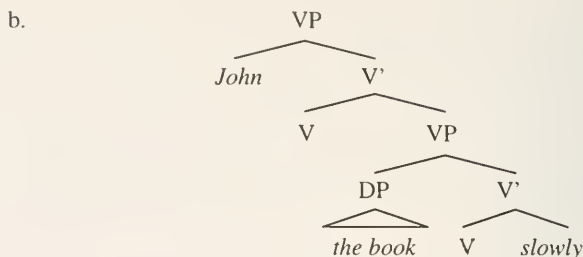
Adverbial adjuncts also played a prominent role in the explosion of functional projections such as INFL (Pollock 1989) in providing a diagnostic for head-positions, and the concomitant existence of functional phrases.

- (3) a. Jean lit souvent un livre
b. *Jean souvent lit un livre
c. John often reads a book
d. *John reads often a book

Adjuncts also figured prominently in Larson's (1988) seminal paper on the shell-structure of VP, where the idea that postverbal adjuncts sit in the most deeply embedded, 'complement' position of the verb was first formulated:

- (4) a. John read the book slowly

* *Caveat lector*. Reviewing a 555-page book is a task doomed to failure. I thought it more useful to situate Ernst's study within a broader theoretical context. I hope that readers of *The Syntax of Adjuncts* will benefit from the present perspective. For important discussions on adjuncts, I thank Norbert Hornstein and Juan Uriagereka.



Both Pollockian and Larsonian insights have influenced in different measures the two (not mutually exclusive) theoretical trends currently dominating the generative enterprise: Kayne's (1994) Antisymmetry hypothesis, and Chomsky's (1995) Minimalist Program. But the fate of adjuncts in each approach has been very different, for reasons that are worth considering, as they shed light on Ernst's contribution.

Although Pollockian reasoning pervaded early minimalism (see Chomsky 1991, 1993; Bobaljik & Jonas 1996), head-movement has become an anomaly in the theory, and Chomsky (2000, 2001a) relegated the process to the PF component, outside the confines of narrow syntax — where minimalist guidelines cease to hold. As for Larsonian shells, they came to be used exclusively for thematic purposes. At various times, Chomsky has noted that 'we still have no good phrase structure theory for such simple matters as attributive adjectives, relative clauses, and adjuncts of many different types' (1995, p. 382, note 22), and that '[t]here has never, to my knowledge, been a really satisfactory theory of adjunction, and to construct one is no slight task' (2001b, p. 15). Chomsky (2000, 2001b) has tried to refine a notion of 'pair merge' for adjunction (as opposed to 'set merge' for complementation), but I concur with Chametzky (2003, p. 206) that

If [Chomsky's Pair-Merge] is what a theory of adjuncts 'that satisfies [the Strong Minimalist Thesis]' looks like, then it seems reasonable to conclude that there is no such thing. Given that the construction behaves 'as if [adjuncts aren't] there apart from semantic interpretation' (Chomsky 2001b, p.15) — that is, that *adjuncts have no syntax* — this is not such a far-fetched conclusion. (emphasis in the original)

Indeed, Chametzky (2000, p. 142) points to adjuncts' core property, their optionality, and argues that

Minimalist theorizing is about trying to discover and use only that which is necessary. There might be, then, a principled tension between the phenomenon of Adjuncts and minimalist theorizing. Adjuncts might be simply the wrong sort of things for there to be a minimalist theory of. Minimalist theorizing aims to employ only operations that cannot be avoided, operating on objects that are mandatory. Adjuncts are not mandatory, and the operation Merge simply assimilated them to the form of mandatory objects (...), and the special nature of Adjuncts entirely disappeared.

Collins (2002) reinforces this point by not mentioning adjuncts at all in his attempt at a genuinely bare phrase structure.

It is in the light of Chomsky's and Chametzky's statements about adjuncts that Ernst's study is best evaluated. But before showing why this is so, I would like to sketch the more optimistic alternative to the syntax of adjuncts that emerged from the pursuit of an antisymmetric approach to grammar.

For theory-internal, mechanical reasons pertaining to the formulation of his Linear Correspondence Axiom (LCA) (if α precedes β , α asymmetrically c-commands β), Kayne (1994) suggested we collapse specifiers and adjuncts. Cinque (1999) built on this proposal to claim that adjuncts occupy the 'specifier' position of abstract, dedicated functional projections. Because of the LCA-driven claim that there can only be one specifier per projection, and the fact that adjuncts iterate, the number of functional projections needed to accommodate the syntax of adjuncts grew exponentially. In so doing, Cinque aimed to capture ordering restrictions among adjuncts, and the parallelism he found between such restrictions and morpheme restrictions inside verbal elements associated with a semantics similar to that of adjuncts. For example, just as verbal elements exhibit a strict ordering, as in (5) (see already Chomsky 1957), adjuncts occur in a corresponding sequence, as in (6), and no other.

(5) He must have been playing in the garden

(6) He allegedly always wrote Martha's letters completely

Cinque suggests we capture the parallelism between (5) and (6) via Baker's (1985) Mirror Principle, which ensures a correspondence between morpheme order and functional hierarchy.

Although Ernst does not directly address the 'Mirror' effects Cinque was interested in, he spends a considerable amount of time providing conceptual and empirical arguments against a Cinque-style approach to the distribution of adverbial adjuncts (see especially his Chapters 3 and 4). Let me briefly mention one set of properties that cast doubt on the antisymmetric take on adjunct.

The issue is simple: adjuncts may (but need not) precede the structures they scope over. The problem arises with a situation whereby a given adjunct follows the structure it modifies, as in (7):

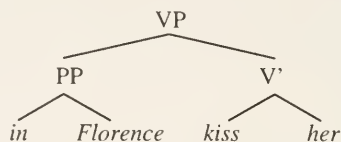
(7) John wanted to *kiss her in Florence*

Semantically we know that *in Florence* modifies *kiss her*, which is standardly characterized in terms of *in Florence* commanding V' . But if *in Florence* commands V' , according to the LCA it should precede V' , contrary to fact. Assuming with Larson (1988) that *in Florence* is a complement of *kiss* won't do. If *in Florence* were a complement, it should be possible to extract from it. This is not the case.

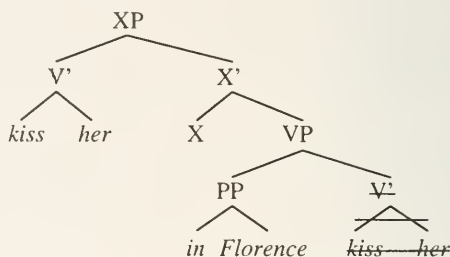
(8) ?* Which country did you meet her in the capital of *t*?
(cf. Which countries did you visit the capital of?)

More recently, Kayne and others have pursued the possibility that sentences like (7) are massively realigned, involving remnant movement of the sort in (9).

(9) Starting derivation:



After remnant movement:



But aside from the fact that it is not always clear what category X really is, and the fact that the displacement of V' goes against the idea that movement is a Last Resort operation, the derivation in (9) makes the wrong prediction that V' is an island, much like standard specifiers (such as subjects). But this is not the case:

(10) Who did you kiss *t* in Florence

Ernst (p. 144f.) is right in saying that

Cinque (...) ha[s] claimed that putting all adverbs into Spec positions, in one-to-one licensing relationships with heads, yields a more restrictive theory than allowing adverbs to be adjoined. It is important to note, however, that this is not the same notion of restrictiveness commonly invoked in the early days of generative grammar, when the child language learner was seen as having to construct the rules of her grammar. (...) [I]n current [Principles-and-Parameters] grammar the child only needs to determine the correct value for [the] parameters provided by UG. (...) Therefore, regardless of whether there are one, or two, or seventeen possible licensing relations, as long as the child does not have to make a choice (...), then this is not a matter of restrictiveness. Rather, it is a matter of simplicity.

One can only agree with Ernst in his statement that 'any putative restrictiveness gained by reducing the number of phrase structural relations should be balanced against increased numbers of movement trigger types, [and] lack of constraints on possible functional heads' (p. 95).

It is fair to say that adjuncts are not specifiers; they don't behave on a par: Adjuncts iterate, specifiers don't; specifiers overwhelmingly surface on the left,

many adjuncts on the right (in SVO languages); some specifiers allow for extraction, adjuncts never do; specifiers license anaphors, adjuncts never do. The list could go on and on. Specifiers and adjuncts just are two different species. Collapsing them structurally loses any hope to distinguish them. Ironically, one of the most compelling arguments in favor of not treating adjuncts and specifiers alike is provided by Cinque himself. (The argument is not discussed by Ernst, who, as I noted above, largely ignores Mirror effects of the type Cinque is interested in.) Cinque (1999, p. 127) notes that 'the order [of functional heads] appears crosslinguistically invariant. The limited cases of apparent variation all seem to involve agreement and negation.' But such cases are not limited at all. Concluding her survey of over 500 genetically unrelated languages, Julien (2000, p. 359) observes that 'there is one inflectional category which does not so easily fit into the ... rigid framework that syntactic analyses [like Cinque 1999] provide. This category is agreement.'

It is remarkable that the main function of specifiers, i.e., agreement, is precisely the one that does not fit the hierarchy that emerges from treating adjuncts as specifiers. It appears that specifiers and adjuncts belong to different systems. Furthermore, it is remarkable that it is an uninterpreted property of the grammar such as agreement that deviates from the rigid order governing the co-occurrence restrictions on adjuncts. It is tempting to conclude based on this that Cinque's hierarchy is a semantic one. (In his conclusion, Cinque (1999) argues that his hierarchy cannot be reduced to semantics, but his arguments are few, and very weak.) But if the relevant restrictions are semantic in nature, it would be redundant to encode them in the syntax.

This is precisely the line of thought that Ernst develops. Ernst argues that for the most part adjuncts freely adjoin to any projection. Adjuncts' possible positions are determined by the interpretive module, which combines adjuncts' lexico-semantic (selectional) requirements (such as scope) and general compositional rules. Semantics is the main filter that acts on the syntax to properly constrain the distribution of adverbials. In addition, considerations on the PF side of the grammar further filter out poor adjunction choices. The two PF considerations Ernst examines in detail are (i) a directionality (/head) parameter, and (ii) prosodic conditions ('morpho-phonological weight'). In short, one can say that as far as adjuncts go, syntax generously proposes, and the interfaces dispose.

After laying out his assumptions and arguing against a Cinque-style approach in Chapters 2-4, Ernst illustrates how the various principles he proposes conspire to yield the correct results for the distribution of post-verbal, VP adjuncts (Chapters 4, 5), vP-adjuncts (Chapter 6), IP-adjuncts (Chapter 7), and CP-adjuncts (Chapter 8). Chapter 9 provides a detailed conclusion, and discusses avenues for further research. The breadth of Ernst's research leaves no doubt that *The Syntax of Adjuncts* will constitute the basis of future work on the distribution and typology of adjuncts. The specifics of the principles that Ernst proposes are bound to be subject to important modifications, but Ernst has shown that they are likely to be the right *kind* of principles for adjuncts.

Let me end by pointing to the biggest problem of all that future work will face regarding the syntax of adjuncts. Ernst has conclusively shown that there is no adjunct-specific syntax, and that the interfaces are responsible for surface patterns. This result compounds the problem of the representation of adjuncts, the operation of adjunction. Current work on Phrase Structure (Chomsky 1994, Collins 2002, Collins & Ura 2001, Chametzky 2000, 2003) emphasizes the role of syntactic relations to deduce properties of phrase markers. But adjuncts lack syntactic relations, they are just 'there' in narrow syntax, and only come in full force in the interpretive components. How this can be integrated within a bare phrase structure theory is the challenge that *The Syntax of Adjuncts* urges us to take part in.

REFERENCES

- BAKER, Mark. 1985. The Mirror Principle and morphosyntactic explanation. *Linguistic Inquiry* 16:373-416.
- BELLETTI, Adriana (ed.). [Forthcoming]. *Structure and Beyond*. Oxford: Oxford University Press.
- BOBALJIK, Jonathan, & Dianne JONAS. 1996. Subject positions and the role of TP. *Linguistic Inquiry* 27:195-236.
- CHAMETZKY, Robert. 2000. *Phrase Structure*. Oxford: Blackwell.
- . 2003. Phrase structure. In: Hendrick (ed.), 192-225.
- CHOMSKY, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- . 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- . 1991. Some notes on economy of derivation and representation. In: Freidin (ed.), 417-54.
- . 1993. A minimalist program for linguistic theory. In: Hale & Keyser, 1-52.
- . 1994. Bare phrase structure. Massachusetts Institute of Technology, MS. [Reproduced in: Webelhuth (ed.), 383-439.]
- . 1995. *The Minimalist Program*. Cambridge, Mass.: MIT Press.
- . 2000. Minimalist inquiries: the framework. In: Martin, Michaels, & Uriagereka (eds.), 89-155.
- . 2001a. Derivation by phase. In: Kenstowicz, 1-50.
- . 2001b. Beyond explanatory adequacy. Massachusetts Institute of Technology, MS. [To appear in: Belletti (ed.).]
- CINQUE, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-linguistic Perspective*. Oxford: Oxford University Press.
- COLLINS, Chris. 2002. Eliminating labels. In: Epstein & Seely, 42-64.
- COLLINS, Chris, & Hiroyuki URA. 2001. Eliminating phrase structure. Cornell University and Kwansai Gakuin University, MS.
- EPSTEIN, Samuel D., & T. Daniel SEELY (eds.). 2002. *Explanation and Derivation in the Minimalist Program*. Oxford: Blackwell.
- FREIDIN, Robert (ed.). 1991. *Principles and Parameters in Comparative Grammar*. Cambridge, Mass.: MIT Press.
- HALE, Ken, & Samuel Jay KEYSER (eds.). 1993. *The View from Building 20*. Cambridge, Mass.: MIT Press.
- HENDRICK, Randall (ed.). 2003. *Minimalist Syntax*. Oxford: Blackwell.

- HUANG, C.-T. James. 1982. Logical relations in Chinese and the theory of grammar. Massachusetts Institute of Technology, Ph.D. dissertation in Linguistics.
- KAYNE, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.
- KENSTOWICZ, Michael (ed.). 2001. *Ken Hale: A Life in Language*. Cambridge, Mass.: MIT Press.
- JULIEN, Marit. 2000. Syntactic heads and word formation: a study of verbal inflection. University of Tromsø, Ph.D. dissertation.
- LARSON, Richard. 1988. On the double object construction. *Linguistic Inquiry* 19:335-91.
- LASNIK, Howard, & Mamoru SAITO. 1984. On the nature of proper government. *Linguistic Inquiry* 15:235-89.
- MARTIN, Roger, David MICHAELS, & Juan URIAGEREKA (eds.). 2000. *Step by Step*. Cambridge, Mass.: MIT Press.
- POLLOCK, Jean-Yves. 1989. Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry* 20:365-425.
- WEBELHUTH, Gert (ed.). 1994. *Government and Binding and the Minimalist Program*. Oxford: Blackwell.



REVIEW

Stockwell, Robert, and Donka Minkova: *English words: History and Structure*. Cambridge: Cambridge University Press, 2001. Pp. 208. Price: (hard-back) \$54.95, ISBN 0-521-79012-3, (paperback) \$19.95, ISBN 0-521-79362-9.

José Ignacio Hualde
University of Illinois at Urbana-Champaign
j-hualde@uiuc.edu

This is a textbook on the structure and origin of, primarily, the learned vocabulary of English. In the introduction, Stockwell & Minkova state that 'if you use a dictionary a lot, you probably don't need this book' (p. 1). Nevertheless, the book would be useful even to frequent users of dictionaries, as it provides information on many topics related to the English lexicon, including the following: possible sources of words, such as compounding, derivation, blending, acronyms, etc., (Chapter 1, 'Word origins'); a brief overview of the history and relations of the English language (Ch. 2, entitled 'The background of English'); the incorporation of foreign words since the Norman conquest (Ch. 3, 'Composition of the Early Modern and Modern English vocabulary'); elements of morphology (Ch. 4, 'Smaller than words: morphemes and types of morpheme'); morphophonological rules in the English lexicon (Ch. 5, 'Allomorphy, phonetics and affixation'; Ch. 6, 'Replacement rules'; Ch. 7, 'Deletion rules and other kinds of allomorphy'; and Ch. 8, 'Fossilized allomorphy: false cognates and other etymological pitfalls'); semantic change (Ch. 9, 'Semantic change and semantic guesswork'); and the stress pattern and other aspects of the pronunciation of unassimilated or incompletely assimilated Latin and Greek words and phases (Ch. 10, 'The pronunciation of classical words in English').

An aspect of this textbook with which I disagree is the abstract or panchronic perspective taken by the authors, where alternations having their origin in a sound change in, say, Old French or pre-Classical Latin, are cast as phonetically motivated rules in the English lexicon. This perspective, which is reminiscent of Chomsky & Halle's *The Sound Pattern of English* and related work, is the source of many descriptive statements and analyses that are bound to cause confusion among non-initiated users of the book. To choose just a couple of examples among many others in the book: on p. 105 the word *democratize* is listed as containing a suffix, *-ize*, 'beginning with a front vowel' which nevertheless fails to trigger T-Lenition. On the next page, we are told that '[f]our distinct changes can take place at the intersection of a morpheme ending in [-t] and a morpheme beginning with [j] or [y]. We can simply get lenition from [t] to [s], as in . . . *importance* . . .' (p. 106). My feeling is that a reader who is not sufficiently familiar with the Chomsky & Halle model and who has understood the difference between spelling

and sounds, which is explained a few pages earlier, will necessarily find all of this mystifying. Sometimes this style of presentation leads to downright anti-historical analyses. For instance, on p. 125, the lexical relation between *table* and *tabulate*, *single* and *singular*, etc., is explained as showing the application of a rule of U-Epenthesis; e.g., *table + ate* → *tabulate* by application of U-Epenthesis. The pronunciation *nucular* for *nuclear* is then presented as a case of optional application where 'U-Epenthesis has yet to achieve full respectability', and *epistolary* (cf. *epistle*) is given as an example 'where the epenthetic vowel is exceptional'. Of course from a historic point of view, this is all incorrect: Latin *tabula* became French *table* by post-tonic vowel deletion, and *tabulate* simply preserves the original vowel; *epistolary* derives from Latin *epistola* (a variant of *epistula*); and *nucular* is an analogical formation (probably based on *secular*, *regular*, *ocular*, and similar examples) which, from a historical point of view, has nothing to do with the other examples cited in this section. The authors may want to argue that, for the benefit of students of modern English, it is more convenient to present these alternations as if they represented a case of epenthesis, but given the fact that the title of this book contains the word 'History', I question the advisability of taking this approach. Students and other readers without previous knowledge may be misled into believing that *tabulate*, *singular*, etc., actually originated from the contracted forms by a process of epenthesis. Again, this is just one example of the general approach adopted in this book.

My preference would have been for the real historical origin of the various alternations to be provided. I believe that most of the time, this sort of historical explanation can be presented without complicating things too much. The benefit to the student who learns — to go back to the example just discussed — that the alternation illustrated by *table/tabulate* has to do with the loss of post-tonic vowels in Gallo-Romance, is that, besides becoming aware of the alternation, she or he will have learned something about historical linguistics. Similarly, to choose another one of the examples mentioned above, as an account of the alternation exemplified by *important/importance*, I would have preferred a concise but accurate description of the sound changes in Late Latin and French that transformed *-antia/-entia* into *-ance/-ence*, rather than a synchronic analysis that makes the questionable assumption that phonologically, English *importance* ends in a front vowel.

In addition to the ten numbered chapters, the book contains two appendices. Appendix I, 'An introduction to dictionaries', is chapter-length and includes a description of different dictionaries and types of dictionaries, providing useful information on this topic. Appendix II, 'Morpheme list', is a list of 427 roots and affixes including, for each of them, its meaning, examples, and its source (mostly Latin and Greek with a few listed as coming from Proto-Indo-European, French, or Old English). This morpheme list is unfortunately marred by certain inconsistencies in the way sources are given. Latin verbs are usually cited in the infinitive, with theme vowel and infinitive ending in parentheses, e.g., L *cant(are)*, but sometimes only the root is given, without any indication that this is not a free form. e.g., L *ambul* 'walk' (for *ambulo*, *ambulare*), L *plec* 'fold, tangle' (for *plico*,

plicare?), and there are also some errors of conjugation, as with incorrect L *sed(are)* 'sit' (correct *sedeo, sedere*). With Latin nouns, sometimes the nominative singular is given, as in L *pars* (for *pars, partis* — we follow the usual convention of citing nominative and genitive singular); other times what we find is the nominative singular with the inflectional ending in parentheses, as in L *lud(us)*; and some other times nonexistent free forms are offered, as in L *oper* 'work, creation' (for *opus, operis*), L *ordin* 'order' (for *ordo, ordinis*), and L *ped* 'foot' (for *pes, pedis*); similarly with adjectives, as with nonexistent L *patient* (for *patiens, patientis*). I think it would have been better to adopt a consistent approach for citing Latin sources. The word *inertia*, listed as an example containing the morpheme *art* 'skill', is listed as having its source in Old French *art*, even though this is an obvious case of direct borrowing from Classical Latin into English. This lack of consistency in the citation of sources makes the usefulness of this appendix somewhat problematic for the student interested in obtaining accurate information on the origin of words.

This book provides much information on a number of aspects of the English lexicon. For instructors who are comfortable with the classical generative perspective this can be a valuable textbook. Other instructors may want to skip some of the chapters and replace them with other materials.

REFERENCE

CHOMSKY, Noam, & MORRIS HALLE. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Studies in the Linguistic Sciences
 University of Illinois - Department of Linguistics
 4080 FLB, 707 South Mathews Avenue
 Urbana, IL 61801

Please send me the following copies of *Studies in the Linguistic Sciences*:

_____	4:1 Spring 1974 Papers in General Linguistics	\$2.00
_____	5:1 Spring 1975 Papers in General Linguistics	\$2.00
_____	6:2 Fall 1976 Papers in General Linguistics	\$2.00
_____	7.2 Fall 1977 Papers in General Linguistics	\$2.00
_____	8.1 Spring 1978 Papers in General Linguistics	\$2.00
_____	8.2 Fall 1978 Papers in General Linguistics	\$2.00
_____	9:1 Spring 1979 Papers in General Linguistics	\$2.00
_____	9.2 Fall 1979 Papers on Relational Grammar and Semantics	\$2.00
_____	10.1 Spring 1980 Papers in General Linguistics	\$2.00
_____	11.1 Spring 1981 Papers in General Linguistics	\$2.00
_____	11.2 Fall 1981 Dimensions of South Asian Linguistics	\$2.00
_____	12.1 Spring 1982 Papers in General Linguistics	\$2.00
_____	12.2 Fall 1982 Papers on Diachronic Syntax: Six Case Studies	\$2.00
_____	13.1 Spring 1983 Papers in General Linguistics	\$2.00
_____	13.2 Fall 1983 Studies in Language Variation: Nonwestern Case Studies	\$2.00
_____	14.1 Spring 1984 Papers in General Linguistics	\$2.00
_____	14.2 Fall 1984 Language in African Culture and Society	\$2.00
_____	15.1 Spring 1985 Papers in General Linguistics	\$2.00
_____	15.2 Fall 1985 Linguistic Studies in Memory of Theodore M. Lighter	\$2.00
_____	16.1 Spring 1986 Papers in General Linguistics	\$2.00
_____	16.2 Fall 1986 Illinois Studies in Korean Linguistics	\$2.00
_____	17.1 Spring 1987 Papers from the 1986 Asian Languages Analysis Roundtable	\$2.00
_____	17.2 Fall 1987 Papers in General Linguistics	\$2.00
_____	18.1 Spring 1988 Papers in General Linguistic	\$2.00
_____	18.2 Fall 1988 Papers in General Linguistics	\$2.00
_____	19.1 Spring 1989 Papers in General Linguistics	\$2.00
_____	19.2 Fall 1989 Contributions of African Linguistics to Linguistic Theory 1	\$2.00
_____	20.1 Spring 1990 Contributions of African Linguistics to Linguistic Theory 2	\$2.00
_____	20.2 Fall 1990 Linguistics for the Nineties, Papers from a Lecture Series in Celebration of the Department's 25th Anniversary	\$5.00
_____	20.3 Meeting Handbook: 13th South Asian Languages Analysis Roundtable, 25-27 May, 1991, University of Illinois at Urbana-Champaign	\$1.00
_____	21.1 Spring 1991 Papers in General Linguistics	\$5.00
_____	21.2 Fall 1991 Illinois Studies in Korean Linguistics, II	\$5.00
_____	22.1 Spring 1992 Papers in General Linguistics	\$5.00
_____	22.2 Fall 1992. Twenty-Five Years of Linguistics at University of Illinois	\$8.00
_____	23.1 Spring 1993. Papers in General Linguistics	
_____	24.1/2 Spring/Fall 1994 (double issue) Proceedings of the 4th FLSMA	\$16.00
_____	25.2 Fall 1995 Language and Gender	\$8.00
_____	26.1/2 Spring/Fall 1996 (double issue) Studies in Chinese Linguistics	\$20.00
_____	29:2 Fall 1999 Forum Lectures, 1999 / Papers in General Linguistics	\$10.00
_____	30:2 Fall 2000 Papers in General Linguistics	\$10.00
_____	31:1 Spring 2001 Diaspora, Identity, and Language Communities	\$15.00
_____	32:1 Spring 2002 Papers in General Linguistics	\$15.00

NOTE: Sales Tax must be included as follows (excluding organizations):

IL 7-1/4%; IN 5%; MI 6%; MN 6.5%; OH 5% and WI 5%.

Checks or money orders in U.S. Dollars, drawn on a U.S. Bank, payable to:
 University of Illinois.

Check or money order enclosed \$ _____ (Please add postage: \$2.00 first issue, \$1.50
 for each additional issue.)

Name: _____

Contributor to *SLS* (25% off)
 Vol. _____, No. _____

_____ e-mail: _____

Visit our

HOME PAGE:

<http://www.linguistics.uiuc.edu/sls.html>

for information on *SLS*,

including:

Editorial policy,

StyleSheet,

Contents of each issue to date,

Complete Author Index,

Ordering information and Order Form

STUDIES IN THE LINGUISTIC SCIENCES
VOLUME 32, NO. 1 (SPRING 2002)

PAPERS IN GENERAL LINGUISTICS

Zsuzsanna Fagyal, Noël Nguyen, and Philippe Boula de Mareüil: From <i>dilation</i> to coarticulation: Is there vowel harmony in French?	1
Liwei Gao: On the nativization of English in China	23
José Ignacio Hualde, Pello Mugarza, and Koldo Zuazo: The accentual system of Mallabia Basque	39
Yong-hun Lee: Resolution algorithms for the Korean reflexive <i>cukicasin</i> : A Categorical Grammar approach	57
Theeraporn Ratitamkul: An alternative analysis to the temporal adverbial <i>when</i> -clause	79

REVIEWS

José Ignacio Hualde: Review of R. M. W. Dixon & Alexandra Y. Aikhenvald: <i>Word: A Cross-Linguistic Typology</i>	91
Cedric Boeckx: Review of Thomas Ernst: <i>The Syntax of Adjuncts</i>	95
José Ignacio Hualde: Review of Robert Stockwell and Donka Minkova: <i>English words: History and Structure</i>	103