Syntax and Variation

Reconciling the Biological and the Social

EDITED BY

Leonie Cornips
Karen P. Corrigan

SYNTAX AND VARIATION

AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE

General Editor

E.F. KONRAD KOERNER
(Zentrum für Allgemeine Sprachwissenschaft, Typologie
und Universalienforschung, Berlin)

Series IV – CURRENT ISSUES IN LINGUISTIC THEORY

Advisory Editorial Board

Volume 265

Leonie Cornips and Karen P. Corrigan (eds)

Syntax and Variation
Reconciling the Biological and the Social

SYNTAX AND VARIATION

RECONCILING THE BIOLOGICAL
AND THE SOCIAL

Edited by

LEONIE CORNIPS
Meertens Institute, Royal Netherlands Academy of Sciences
KAREN P. CORRIGAN
University of Newcastle

∞ ™ The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences — Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Table of contents

Toward an integrated approach to syntactic variation

A retrospective and prospective synopsis

Leonie Cornips and Karen Corrigan

Meertens Institute, Royal Netherlands Academy of Sciences / University of Newcastle, UK

1.   Overview

1.1   Stimulus for the volume and its overarching aim

Five of the contributions (Adger & Smith, Barbiers, Cheshire, Gervain and Zemplén as well as Henry) arose from invited presentations at the workshop on Syntactic Variation organised by the editors of the present volume that was held in June 2003 during the *Second International Conference On Language Variation In Europe* (ICLaVE 2) at the University of Uppsala. The primary aim of this workshop was to initiate cooperation between internationally renowned generative and variationist linguists with a view to developing an innovative and more cohesive approach to syntactic variation. The present volume then evolved by inviting further contributions from like-minded scholars so that the work as a whole would contain treatments incorporating the analysis of external factors into accounts focusing on the internal linguistic conditioning of syntactic variation and change cross-linguistically.

We have partitioned the book into four major parts, grouping chapters that have orientations in common together. Part I, which contains the contributions by Cheshire, Muysken and Sorace, focuses on the locus of syntactic variation and aspects of modularity. The chapters in Part II by Henry and Gervain and Zemplén are oriented towards methodological innovation with an emphasis on personal pattern variation.[1] The contributions in Part III by Adger and Smith, King and Van Gelderen seek to address syntactic variability in real and appar-

ent time with particular emphasis on the extralinguistic factors of age, gender and style. Finally, Part IV, which consists of contributions on Dutch (Barbiers) and Romance (Benincà & Poletto) is devoted to synchronic variation across geographical space.

1.2 Wider context

We are not the first to point out that researchers who espouse the frameworks encapsulated by the umbrella terms 'biolinguistics' and 'sociolinguistics', diverge quite rigidly in terms of both their methodological approaches and theoretical persuasions.[2] Although there remain certain formal resonances between the paradigms since the early days of their inception, the fundamental differences between them created a schism that has persisted through most of the later twentieth and early twenty-first centuries (cf. Cheshire 1987, 1999; Cornips & Corrigan 2005; Hudson 1995; Henry this volume; Kroch 1989; and Sankoff 1988a). In this regard, Wilson and Henry (1998:2) note that "there have been few real attempts to marry these seemingly divergent positions" and Meechan and Foley (1994:63), likewise, suggest that theoretical syntacticians and sociolinguists "rarely, if ever, cross paths".

The welcome relaxation of the generative position on the status of externalist accounts from that of Smith (1989), typified in the quotation below from Chomsky (1999:34), demonstrates that the time may well be ripe for a more integrated approach such as those introduced in Cornips and Corrigan (2005); Henry (2002); Meechan and Foley (1994); Meyerhoff (2000); Van der Wurff (2000) and attempted more exhaustively in the present volume.

> Internalist biolinguistic inquiry does not, of course, question the legitimacy of other approaches to language, any more than internalist inquiry into bee communication invalidates the study of how the relevant internal organization of bees enters into their social structure. The investigations do not conflict; they are mutually supportive. In the case of humans, though not other organisms, the issues are subject to controversy, often impassioned, and needless.

As such, entertaining "Reconciling the biological and the social" could well be described in Kuhn's (1970) terms as the initial phase in the creation of a mature scientific community, united by a single paradigm. It is hoped that by doing so we will overturn the present situation which still fits all too squarely within Masterman's (1974:74) diagnosis of the problems engendered by an "immature science" (Kuhn 1970:182):

Each sub-field as defined by its technique is so obviously more trivial and narrow than the field as defined by intuition, and also the various operational definitions given by the techniques are so grossly discordant with one another, that discussion of fundamentals remains, and long-run progress (as opposed to local progress) fails to occur.

1.3   The acquisition of local and supralocal varieties

Not surprisingly, given its orientation, variationist sociolinguistics often focuses on speakers of local varieties or dialects (cf. the papers by Adger & Smith, Cheshire, Henry, King and Muysken). As the contributions to this volume by Barbiers, Benincà and Poletto, Gervain and Zemplén and Henry confirm, the necessity for generative researchers with interests in syntactic microvariation to also attend to these vernaculars seems to be on the increase.[3] Two separate, but related, questions arise with respect to this kind of data that we feel should be addressed in this introduction since they have important implications for our discussion of the major themes of the volume in the sections which follow, namely: (1) how are non-standard varieties acquired by the individual/community? and (2) to what extent are such vernaculars subject to variation and change within individuals/social groups and across diachronic, diatopic and diatypic dimensions?[4]

The field of biolinguistics envisages linguistic change as primarily being driven by the acquisition process because learners have the option of adopting innovative settings for the parameters provided by Universal Grammar. There has, however, never been the same emphasis on acquisition within the sociolinguistic paradigm. Indeed, Roberts (2002:333) has recently claimed that investigations into "the acquisition of variable features by young children" are "relatively new." Early studies in this model, such as Labov's (1989) investigation of (-*t, d*) deletion in Philadelphia, did find that pre-pubescent language learners acquired the socially situated variability that characterized their parents' speech patterns. Nevertheless, there are several reasons why this age group has been relatively neglected in sociolinguistics by comparison to their importance as data subjects in the evolution of the biolinguistic paradigm. In the first place, discriminating between developmental errors and genuine variability can be highly problematic. Secondly, the methodological practices favoured by sociolinguists which require large subject groups and many hours of data do not easily lend themselves to the recording of very young children. Some may be taciturn in the presence of adults and even their peer group, they may favour telegraphic speech and tend to have short concentration spans, all of which

make it very difficult to gather enough variable data of the right kinds. That there is much to be gained by attempting to overcome these problems is clear from the findings of Chambers (1992), Roberts and Labov (1995) and Foulkes et al. (1999) which examine variable caretaker input and child output, though their investigations focus solely on the phonological component. Extending this research "above and beyond phonology" (Sankoff 1980) to explore *grammatical* variability in this age group will be one of the most interesting future challenges of the integrated approach to syntactic variation advocated here.

A related issue, of course, is whether we consider adult speakers in contemporary western communities in particular to be 'true' monolinguals anyway, given the social milieu that generally pertains (highlighted, for example, in the 2004 collection entitled *The Acquisition of Sociolinguistic Competence*, edited by Bayley and Regan). Thus, in addition to the fact that every dialect is naturally a heterogeneous system, varieties rarely exist nowadays in absolute isolation. Indeed, most competent speakers of language X can usually resort to a range of varieties along a continuum from standard to non-standard, depending on social and discourse contexts. Indeed, while syntax is often viewed within sociolinguistics as a marker of cohesion in large geographical areas, syntactic variants may also act as marker of local identity, as is the case with variability in the phonological component (cf. Cornips in press). It is surely not beyond the pale, therefore, for sociolinguists to claim that dialect systems of even adult speakers are not static but are participating in ongoing processes of change as a result of social, political, cultural and economic influences. Even in those increasingly rare communities in which supralocal models are absent, face-to-face interactions are often polylectal (cf. Auer ms. and Harris 1985). Indeed, as Henry, this volume, demonstrates more attention should be paid by both sociolinguists and biolinguists to the phenomenon of idiolectal variation in this regard. In addition, dialect speakers may be raised 'bilingually' from birth in the local dialect and a supralocal variety, thus being, in effect, simultaneous L1 acquirers. It is also possible that in certain linguistic contexts (reminiscent of diglossic situations) children acquire the supralocal variety at a somewhat later age and, as such, should be considered early child L2 acquirers of the standard. As a result, it is likely (following the views expressed by many of the contributors to Bayley & Regan 2004 and in Sorace this volume, regarding adult L2 acquirers of typologically different languages/varieties) that in so-called non-standard communities, the total exposure to both the dialect and the standard variety is reduced compared to that of monolingual standard or dialect speakers and this is worth bearing in mind.

2.   Outline of contributions and their methodologies

There are a number of points of synergy and contrast with respect to the methodologies adopted and the grammatical phenomena analysed in the present volume and it is to these issues that we now turn. As regards the data-sets mined, for example, with the exceptions of King's paper on Acadian French, Muysken's investigation of Ecuadorian Spanish and Van Gelderen's analysis of the *Corpus of Spoken Professional American English* (*CSE*), contributions focus on the linguistic landscape of Europe. In addition, the data described *passim* is predominantly spoken and synchronic, although Van Gelderen's paper on the history of English, which draws on written sources, is unique in both these respects. However, the approach of all the authors to their materials is a comparative one, drawing on parallels and distinctions between: (i) idiolects (Henry's investigation of expletive *there* agreement and Gervain & Zemplén's investigation into focus-raising across divergent Hungarian lects); (ii) dialects (like Barbiers' account of word order strategies in Dutch varieties and Benincà & Poletto's treatment of agreement and person features in Romance) and (iii) languages (as is the case with Sorace's contribution). Adger and Smith, Cheshire, Henry and Van Gelderen all offer accounts of different dialects of English in the British Isles and North America, including standard varieties, though they differ with respect to the manner in which the data was collected and subsequently mined. The papers by Adger and Smith and Cheshire, for example, focus on samples of tape-recorded speech using the classical sociolinguistic interview method and they, therefore, share the approach of King and Muysken in this regard. Van Gelderen's paper, by contrast, is corpus-based using data from different periods, dialects and styles of English. Henry's account of another variety of English (those vernaculars spoken in Northern Ireland) is more akin to the classical generative method since her analysis relies on accessing intuitions. As such, it is similar to the oral and written elicitations described in the work on Dutch by Barbiers, on Hungarian by Gervain and Zemplén and that of Benincà and Poletto and Sorace on members of the Romance language family. However, there are particularly innovative aspects of the methodologies described by both Henry and Gervain and Zemplén that are worth foregrounding here. Henry's method is unique in that it relies on a predetermined set of test sentences only in the initial phase of data collection, the bulk of her intuitive data coming from long term discussions of acceptability judgements with individual native speakers. Likewise, Gervain and Zemplén's contribution is unusual in the context of theoretical treatments of syntactic variation in that it takes a quantitative approach (akin

to mainstream sociolinguistics) in its reporting of cluster analysis results for the acceptability judgements of individual native speakers.

As far as syntactic phenomena are concerned, the features addressed in this volume range from treatments attending to aspects of the DP (such as Benincà & Poletto's and Van Gelderen's innovative accounts of the pronominal systems of Romance and English, respectively) to novel analyses of word order strategies (like Barbiers' contribution on verbal clusters in Dutch and Gervain & Zemplén's account of variation in the constructions amenable to focus-raising in Hungarian). In between these two poles, there are integrated accounts of various syntactic features that have often been addressed rather less successfully in the past by researchers working independently within either the biolinguistic or sociolinguistic paradigms. Thus, Sorace's and Van Gelderen's contributions address the popular generative topics of parameter setting/re-setting as well as pro-drop and verb second phenomena *inter alia*. The unusual patterning of verbal agreement, *do*-periphrasis and negation in Buckie English is tackled by Adger and Smith while Cheshire's paper revisits conventional sociolinguistic accounts of variation in pronoun tagging and verbal *-s* and introduces the relatively unrecognised phenomenon within this paradigm of independent adverbial clauses. Issues surrounding variation in the systems of inflection and case feature prominently in the papers by Gervain and Zemplén as well as Van Gelderen, who also posits a potential relationship between co-ordination and different classes of noun. IP phenomena, as previously mentioned, are addressed by Barbier's contribution and Muysken's paper gives a detailed account of the gerund in Ecuadorian Spanish. The latter is permitted in a variety of constructions (as it is in English) and these gerundial expressions are especially interesting from our perspective since they are subject to very considerable internal and external variation. If one were forced to isolate the single most prevalent grammatical phenomenon discussed in the volume it would have to be agreement. Although it is the focus of the contributions on expletive *there* by Henry and negative concord by Adger and Smith as well as King, it also features in some form or other in almost all of the papers. This is perhaps not unexpected given the fact that this variable is relatively immune from some of the problems that often beset the application of traditional Labovian methods (originally developed for socio-phonetic/phonological analysis) to the syntactic component (see Cheshire this volume and Cornips & Corrigan 2005). Moreover, considerable generative research effort has been invested in the analysis of the internal structure of IP (see Bobalijk & Jonas 1996; Pollock 1989; Rizzi 1997 *inter alia*), with the result that the facts of verbal agreement, for instance (see Chomsky 1995, 2000, 2001b and Pesetsky & Torrego 2004) are

much better understood than other sites of syntactic variation such as the pre-fabricated expressions described in Cheshire's chapter in the present volume.

3. Major themes addressed

3.1 An integrated theory of syntactic variation

As noted previously, this volume offers a range of papers situated within two of the most salient current frameworks for analysing syntactic variation and change between and within language varieties. Common ground for all the papers is that each attempts to achieve an adequate understanding of the mechanisms determining syntactic variation and change by combining insights from both paradigms. Many researchers have previously claimed that such a bridge is a prerequisite to enable us "to understand language variation and change as they are driven by social factors but constrained (at one level) by the nature of possible grammars" (Wilson & Henry 1998: 8 and see Sells et al. 1996b: 173). Achieving this largely depends on how much variationist and generative researchers are actually willing to countenance and accommodate viewpoints from both disciplines. In our view, the variationist approach would benefit considerably (as King this volume, also argues) from elucidating "microvariation by analysing very closely-related grammatical systems using the technical apparatus that the generative tradition makes available." The more 'classic' variationist contributions here are, therefore, innovative in just this respect in that they use formal insights from generative theory (Muysken), and Minimalism, in particular (Adger & Smith, Henry and King) to explicate patterns of variation and change. On the other hand, we believe that the generative approach has much to gain from a perspective in which the organization of the grammar may be seen as somehow reflected in patterns of usage (Taylor 1994; Van der Wurff 2000) and by availing of a variationist methodology, one can then truly catch "a glimpse of grammatical structure" (Meechan & Foley 1994: 82; Sells et al. 1996a: 624). Thus, quantitative results may not only lend strong support to structural analysis (Pintzuk 1995; Van der Wurff 2000) but they also provide more evidence for microvariation between closely related grammatical systems exhibiting 'orderly heterogeneity' that can, in turn, be correlated with external variables of one sort or another. Far from side-stepping the fact that syntactic innovations propagate at different speeds diachronically, diatopically and diatypically (as most 'classic' generative studies are wont to) those who work primarily within this paradigm and are repre-

sented in this volume (Barbiers, for example) understand the importance of considering quantitative differences to be evidence for aspects of their theoretical analysis. Gervain and Zemplén, especially, have applied the Principles and Parameters framework to Hungarian focus-raising and their findings with respect to native speaker judgements regarding this phenomenon have led them to a similar view to that expressed in Henry's contribution, i.e. that variation is a ubiquitous problem for the generative linguistic enterprise as it is currently conceived. It would seem that unless the framework can devise a systematic treatment of varying intuitions in terms of both their collection and classification, coupled with some mechanism for incorporating quantitative methods, then certain syntactic phenomena will forever remain elusive. The importance of the variationist approach in this regard is also highlighted in both the study by Muysken of the gerund in a partly bilingual Quechua-Spanish community in Andean Ecuador and in Benincà and Poletto's account of variable person features in Romance dialects. Each of these contributions clearly demonstrates that there are implicational relationships between their sets of data in terms of frequency and probabilities that should not be ignored. Muysken, for example, claims that the frequent use of the gerund in one construction appears to be linked to similarly frequent usage of the same variable in other constructions. Likewise, Benincà and Poletto claim that morphological extension is a probabilistic phenomenon, i.e., the more features which two forms have in common, the more probable extension there will be. As such, we strongly agree with the view of Sells et al. (1996b: 173) articulated below:

> Variation theory needs grammatical theory because a satisfactory grammatical characterization of a variable is a pre-requisite to decisions about what to count and how to count it, and it is an essential element in the larger question about where variation is located in speakers' grammars.

The contributions in this volume by Adger and Smith, Barbiers, Benincà and Poletto, Gervain and Zemplén, King and Van Gelderen are testament to the fact that for the generative enterprise, the inclusion of quantitative analyses of usage patterns is critical since they provide insight into the categorical or variable behaviour of the variants in question. Likewise, the chapters by Adger and Smith, Cheshire, Henry, King and Muysken robustly demonstrate that variationist sociolinguists who resort to formal linguistic theory can find novel and more effective measures for deciding which variants are syntactically related and which are syntactically remote. This suggestion was already tentatively made by Corrigan (1997: 224–227, 2000b) and Wilson and Henry (1998: 11) in their analyses of constructions such as (1a) and (1b) below and we are pleased

to be able to incorporate in this volume Henry's most recent findings regarding this phenomenon amongst speakers of Northern Irish English vernaculars.[5] (1a) has a preverbal subject and singular agreement whereas (1b) is an expletive *there*-construction showing singular agreement with a postverbal subject (DP-associate):

(1) a. When **the grapes was** in season
"When the grapes were in season."
(Corrigan 1997:215, U1455/L2399-2400/1945F/MS1112)

   b. There *was* **two priests** lived there
"There were two priests who lived there."
(Corrigan 1997:218, U132/L212-213/1973M/MS1810)

Within the variationist paradigm, agreement phenomena of this kind naturally appear to represent syntactic variants of one and the same linguistic variable (see Cheshire this volume; Eisikovits 1991; Hazen 1996; Meechan & Foley 1994 and Poplack & Tagliamonte 1989) whereas they would be considered syntactically remote in generative syntax on account of the difference in their formal syntactic behaviour (see Belletti 1988; Chomsky 1991; Corrigan to appear; Henry 2002 this volume; Roberts 1997 and Wilson & Henry 1998).

3.1.1 *Questions to be addressed in an integrated theory of grammar*

Bearing issues such as these in mind, we would like to argue that the approach taken by contributors to the present volume finally makes it possible to address fundamental questions such as: (1) Why is grammatical differentiation non-arbitrary, bounded and predictable (cf. Cornips 1998)? and (2) Why is the same degree of variability not shared by all individual speakers despite the fact that 'orderly heterogeneity' can be discerned across the community. From this perspective, Barbier's contribution contends that:

> Generative linguistics and sociolinguistics are complementary in that it is the task of sociolinguistics to describe and explain the patterns of variation that occur within a linguistic community, given the theoretical limits of this variation uncovered by generative linguistics.

Moreover, the chapters by Gervain and Zemplén and Henry provide strong evidence to support the claims of Cornips and Poletto (2005) that linguists should strive towards a more systematic collection strategy for eliciting intuitions in 'spontaneous' and experimental elicitation settings, particularly given the open-ended nature of syntax. Heretofore, spontaneous data within the variationist paradigm has been considered to be far more authentic than in-

tuitions or elicited data of the sort advocated in some of the contributions to this volume (cf. Coupland 2003 and Rickford 1987).[6] The request for progress with respect to data collection techniques is also crucially addressed to generative theorists who, as their research proceeded in the later twentieth and early twenty-first centuries, have relied on data that has become increasingly subtle and, in many ways, therefore, more challenging (Gervain 2002; Labov 1996 and Schütze 1996). However, as Gervain (2002) has recently argued (and see Henry in press), we feel that it is important to note that degrees of acceptability or grammaticality uncovered by biolinguistic methods are not in themselves problematic. Rather, what seems most controversial to us is the imprecise manner in which they are accommodated within generative analyses. It is our view that attention to these issues will, for example, clarify what scale, if any, is being used, how the different degrees of intuition relate to one another and how they should be interpreted in the analysis (cf. Gervain 2002).

3.1.2 *The locus of syntactic variation*

A persistent problem addressed by many of our contributors relates to the manner in which a well-known truism of variationist theory is accounted for, i.e. that individual speakers can use several variants of the syntactic variable (when maintaining the same style level). In fact, this issue is related to questions posed by successive generative models concerning the locus of syntactic variation, its restrictions and predictability. In the literature, several alternative approaches to this 'choice' are suggested, as outlined by Muysken (this volume). Three options with respect to 'variability' are offered here: (1) it is placed outside the grammatical mechanisms (cf. the contributions to this volume by Adger & Smith as well as Cheshire and King); (2) it is located inside the grammar by re-introducing optional rules (Henry 2002, this volume and Wilson & Henry 1998) and, finally (3) it is brought about by movement constrained by agreement (Barbiers this volume). The first option was originally advocated by Kroch (1989) who claimed that the grammar was a blind, autonomous system and the notion of 'choice' (optionality, variability) was not part of it. Instead, the individual speaker avails of separate or competing grammars when expressing variability. Adger and Smith (this volume) likewise, argue that the notion of 'choice' cannot be accounted for within the autonomous grammar. However, in contrast with Kroch's vision, this doesn't imply that individual speakers "have different grammars, per se, but rather a range of lexical items open to them, some of which will have syntactic effects." In their analysis, the notion of 'choice' concerns the level that serves as the input for the autonomous, grammatical system. Henry's contribution to this volume on the other hand reflects

her (1995, 2002) position that individual grammars include variability and, consequently, that the speaker has a real choice in terms of syntactic operations like optional verb movement and agreement, for instance. Different again is Barbier's claim in this volume that not all variation can be reduced to morphosyntactic or spell-out properties but that different dialects may share the same grammar. Such a stance, therefore, permits (indeed predicts) a certain degree of optionality (i.e. variation is thus taken to be an inherent property of the grammatical system).

Only when a suitably mature theory of syntactic variation has evolved will linguists be in a position to adequately address such questions. This is particularly so in the current generative research climate in which Minimalism stoically considers the grammatical system to be autonomous and variation is permitted to occur only at the moment of performance, i.e. when this endowed system is put to use.[7] What is particularly encouraging about the modular approach introduced in Part I of the volume, is that the model allows for a meticulous examination of the extent to which variation is part of the grammatical mechanisms employed and where exactly performance fits (both on the level of proposing hypotheses and evaluating the data). Critical too is the framework's methods for determining whether or not the range of syntactic variation is the same or different in kind, therefore, neatly side-stepping the taxing issues raised in §§3.1–3.1.2 above.

3.1.2.1 *Future avenues in socio-syntactic research: Interface levels?* Three authors in this volume (Cheshire, Muysken and Sorace) are the most vociferous with respect to the necessity of a modular approach, though it is a latent theme in other chapters too. Muysken, in particular, argues in his contribution that this orientation is needed to explain the range of syntactic variation encountered in natural languages. He discusses the (over)use of the Spanish gerund amongst Quechua-Spanish bilinguals and demonstrates how this feature interacts with the cognition, interaction, semiotic and syntactic modules that are assumed to comprise our linguistic competence.[8] It is clear from his analysis that some properties do not unambiguously fall into a single module.[9] Indeed, many authors in this volume consider variation to be engendered by the interaction between the syntactic component and other modules of our linguistic competence. Adger and Smith, for example, argue that the source governing the choice of which lexical item to enter into the syntactic component is influenced by (amongst others): (1) processing e.g. the ease of lexical access and (2) the 'interaction' module, i.e. optionality hinges on speaker-hearer relationships, and on notions of social identity. Similarly, Benincà and Poletto in

their contribution argue that the morphological and semantics modules are jointly responsible for the processes of analogy that they uncover in dialects of Romance.

Sorace's chapter explores these ideas from a language acquisition perspective arguing that features at the interface of syntax and discourse (her interpretation of Muysken's interaction module) display 'emerging' variability (or 'optionality' in her terms). Her paper explicitly addresses the question as to whether interface problems are internal to the learner's representation of syntactic knowledge, or are external to these representations, being created instead by computational difficulties with respect to integrating knowledge from different domains.

Finally, Cheshire's contribution to the volume explicitly addresses the interaction module, which she argues is directly responsible for the use of prefabricated expressions and variants expressing affective meanings in spoken language. At first sight these phenomena may appear to be instances of syntactic variation but she makes a strong case to suggest that they are, in fact, the result of interactive capacities that are responsible for the management of the sequential nature of information exchange, and for the cohesion of human discourse. Only at this interface level, can a distinction be made between pragmatic variation (communicative intent) and syntactic variation (equivalent constructions).

Of course it still remains to be seen whether a modular approach can provide an answer to the question posed earlier with regard to the extent to which variation is an inherent property of the grammatical system. According to Sorace, experimental research on native speakers points to a distinction between violations of 'soft' constraints, which trigger gradient linguistic judgements, and violations of 'hard' constraints, leading to categorical judgements. She draws the important conclusion that hard constraints are purely syntactic in nature, thus, brought about by the syntax module only and that these do not play a role outside language proper, as formulated by Muysken (this volume). Soft constraints on the other hand tend to be associated with the mapping between syntax and other modules such as lexical semantics, pragmatics and information structure. The latter two are defined by Muysken, for example, as belonging to the interaction module and it is interesting that Sorace's experimental results in another context provide further evidence supporting his view that variation emerges in the interface of the syntax and interaction module rather than in the syntax module alone. Falling out from this perspective, is Muysken's opposition to the view that properties which are normally formulated within the generative framework as 'principles' within the syntax, such as

endocentricity or headedness and the effects of 'movement', are indeed located here. By contrast, Muysken argues that such notions of hierarchy probably belong wholly to the cognition module. Moreover, it is likely that other capacities fall within this domain, such as the concept of (extended) projections. The findings of Sorace are, therefore, doubly important since they may indicate that the syntactic capacities alone could be responsible for strict boundaries between language varieties due to the categorical judgements which she reveals.

The chapter by Gervain and Zemplén is also illuminating in this respect. They report that, in the elicitation task discussed in their study, individual speakers of Hungarian display only hard constraints, i.e. categorical judgements with respect to a specific movement strategy like focus-raising (their group 3 doesn't allow focus-raising at all, whereas groups 1 and 2 do). Interestingly, the individual speakers who do allow focus-raising display soft constraints, namely, gradual judgements concerning the case of the focussed constituent and agreement of the embedded verb, which echo the sort of constraints revealed by Henry's contribution (also in Part II), though for a different variable.

Moreover, Sorace's chapter crucially shows a remarkable convergence among data from different domains. Hence, speakers displaying L1 attrition and those who are near-native L2 speakers present similar patterns of stability and variation. These outcomes within an essentially generativist account are critical since they are clearly reminiscent of sociolinguistic findings in which variation in obsolescing forms appears subject to the same independent linguistic constraints as those uncovered in 'healthy' languages and dialects (see King 1989 and Wolfram & Schilling-Estes 1995). In Sorace's case, constructions that belong to the syntax proper are fully acquired in L2 acquisition and are retained in L1 attrition. In the bilingual language contact situation that Muysken reports on, it is interesting to note that the over-use of gerunds as main verbs by the so-called *cargadores,* i.e. non-local rural monolingual Quechua-speaking Indians, has nothing to do with their native language as such. Instead, it is a reflection of the general learning strategies that Sorace also discusses.[10]

Importantly, Sorace argues that the near-native speakers of Italian, mentioned above, do not lack syntactic knowledge since they have actually acquired a null subject grammar. What seems to be at stake is their knowledge of the appropriate felicity conditions for the proper use of overt and preverbal subjects. The optionality in their grammar is at the level of the discourse, relating to the distribution of pronominals and the placement of subjects.[11]

The implications of the modular approach for an integrated theory of syntactic variation is that the locus of the phenomenon is most likely to be

where the syntax module is mapped to other domains. Sorace suggests that the conditioning of the syntactic variants as categorical is linked to the syntax module. However, the conditioning of the syntactic variants as variable is probably due to the combination of underspecified features and the mapping between the syntax module with discourse/pragmatic knowledge (interaction module). Moreover, Muysken convincingly demonstrates that the range of syntactic variation is not the same everywhere. In other words, it is not brought about by the syntax module alone but involves the interplay between it and the cognition, interaction and semiotic modules. What is relevant for sociolinguists is that the over-use of a syntactic variant may be facilitated by the semiotic module in that the variant (the gerund in Muysken's case study) functions as a linguistic sign indexical of external factors such as age, gender, social class or ethnic group membership.

Thus, the interface between the syntax module and the domains of discourse/pragmatics, (i.e. the interaction module where 'soft' constraints are violated) is the arena in which variation will be more frequent or emerge more easily than would be the case in the syntax module alone. From this, we may put forward the hypothesis that variation that has social meaning is more frequently located in this domain than it is in others.[12]

All that being said, we agree with the opinions expressed by Cheshire, Muysken and Sorace in their contributions to the present volume that there is much current progress in the field of syntactic variation (optionality) but there is no comprehensive explanation of the facts they unveil in their investigations currently available. Paying closer attention to modularity and interface levels may well prove critical to enhancing our understanding of the locus of variation on which these issues hinge.

3.2 External and internal constraints on syntactic variation

Having considered the major themes of the volume with respect to the grammatical system we finally turn to review issues addressed by our contributors relating principally to the embedding of syntactic variants within geographical, social and other external matrices but also to the effects of internal linguistic constraints.

A range of factors known to correlate with linguistic variation (see Chambers et al. 2002) are addressed in the present volume, including: (1) The social dynamics of syntactic and pragmatic variants which are addressed by Adger and Smith, Cheshire, King and Muysken; (2) Syntactic change in real time which is the focus of Van Gelderen's analysis of data from various histor-

ical English sources; (3) Stylistic factors which are tackled in rather different ways by both Cheshire and Van Gelderen; (4) Regional variation described in the chapters by Barbiers and Benincà and Poletto and (5) Idiolectal or personal pattern variation which is the focus of Henry's and Gervain and Zemplén's contributions.[13]

The chapter on Buckie English by Adger and Smith, as noted in §2, explores *was* versus *were* agreement and *Do* absence/presence in negative declaratives. It uses classical Labovian methodology with respect to quantifying the variation across its entire range and isolating categorical versus variable behaviour amongst different groups of speakers. Another approach is illustrated in the papers by Cheshire, King, Gervain and Zemplén, Muysken, Sorace and Van Gelderen, each of which considers the frequency of use of syntactic variants in different contexts.

There are interesting congruences and disparities with respect to the contributions by Adger and Smith and King regarding the dynamics of social variation. Both chapters focus on patterns of verbal agreement, albeit in different linguistic communities (the marginalised English of Buckie, North Eastern Scotland and the French spoken in Newfoundland in Atlantic Canada, respectively) and the models underpinning both their analyses assume that lexical items are simplified in that they lack any phonological information. They are, in effect, just bundles of syntactic and semantic features which are spelled out as morphemes at some point in the derivation (Halle & Marantz 1993). Interestingly, their results demonstrate that verbal agreement phenomena have been appropriated rather differently by speakers within these communities. In the Buckie data, there is a slight tendency for *was/were* to correlate with age but not gender whereas no such correspondences are attested in King's investigation of agreement phenomena in Acadian French. This is the case too with the second variable examined by Adger and Smith, i.e. *Do* absence in negative declaratives, which appears not to be undergoing change across generations nor to be the marker of gender differences, for example. In this respect, therefore, the findings of both papers can be distinguished from those of Cheshire, this volume, as regards so-called 'lone' *when* clauses. She finds these predominating in male as opposed to female narratives where they function as a marker of shared reminiscence, a narrative style not characteristic of the rather more monologic narratives preferred by females.

Although, Muysken's investigation of the gerund differs from those of Adger and Smith and King in a number of respects (particularly the possibility that at least some of the effects reported therein may be the result of substral influence from Quechua), a number of his objectives with respect to exploring

the social trajectories of linguistic change in Ecuador are similar. Thus, applying the same multivariate analysis techniques of VARBRUL (Sankoff 1988b) that King used in her investigation, revealed that age, gender, and educational background did not have a significant effect on the use of gerunds amongst Ecuadorean Spanish speakers. By contrast, those who were Quechua-dominant bilinguals showed an increasing propensity for using gerunds by comparison to both Spanish monolinguals and Spanish-dominant bilinguals. In addition, Muysken's contribution is the only paper in the volume where the stalwart sociolinguistic variable of class is an important focus (though Cheshire's analysis of pronoun tags used by adolescents in Hull, England incorporates some discussion of the variant as a class marker). Muysken's findings suggest that the most disadvantaged social group (the *cargadores* noted above) had higher frequencies of gerund usage than the *gente*, who represent the local elite.

Henry's paper, likewise, can be singled out with respect to the external factor at the centre of her account, since her aim is to focus on idiolectal variation with respect to grammaticality judgements pertaining to expletive *there* constructions in Northern Irish English. In a sense then, her research seeks to address more formally the importance of accommodating what Chambers (2003:93ff.), in a sociolinguistic context, has recently termed "Oddballs" and "Insiders", i.e. individuations that upon closer inspection may turn out to have social and linguistic significance, though these have largely been ignored by sociolinguists in favour of group norms. The importance of the findings reported in Henry's chapter relates to the extent to which personal pattern variation with respect to a tiny subset of the grammatical component exists. This new evidence requires an adequate explanation from both the generative and variationist paradigms. Not only does it run counter to views in the former that variability at this microscopic level actually exists, Henry's chapter also demonstrates that it is worth accommodating in a sociolinguistic framework that has become overly-concerned with meeting the needs of a bell-curve approach to societal difference.

Aspects of the contributions by Adger and Smith, Cheshire, Henry, King and Muysken are reflected in the paper by Van Gelderen, who similarly attends to the potential for internal linguistic constraints to operate on the syntactic variable, although she is dealing with diachronic as opposed to synchronic data. Thus, Adger and Smith note that grammatical person strongly affects the choice between standard and non-standard variants in their study. Similarly, King records the conditioning of verbal agreement in Acadian French (specifically in subject relative clauses versus other types of clause, involving the C-domain and left periphery) and Muysken finds that the non-standard

gerund variant was considerably more frequent: (1) when the adverbial clause followed the main verb than when it preceded it and (2) when the subjects were identical. Although the picture of inherent variability that emerges from Henry's account is much less clear-cut (though this is not unexpected given her methodology and orientation), she also finds that the variation is linguistically constrained in certain respects. It can, for example, be construction specific (speakers can favour a lack of concord in expletive structures but disfavour it in other contexts) and it is subject to other "natural processes" demonstrated cross-linguistically such as 'position within the clause', 'animacy' and 'quantity'.

In a similar vein, Van Gelderen describes internal variation in the nominal and pronominal systems of English from various periods. She explores, for instance, whether pronouns are more or less likely to be coordinated than nouns due to a language universal, the 'Head Preference Economy' principle, i.e. "if possible, be a head, rather than a phrase." Echoing some of the ideas introduced in §§3.1–3.2 above, Van Gelderen assumes an interface with the discourse domain such that old or given information, and pronouns typically occur as sentence initial items, and new, or focussed information, i.e. nouns, comes towards the end. Consequently, subject functions are less frequent with nouns than non-subject functions and pronouns are naturally used more often with subject functions. Her analysis of the *Helsinki* (*HC*), *British National* (*BNC*), *CSE* and Shakespeare's *First Folio* electronic corpora demonstrate that this pattern has been sustained in English since Anglo-Saxon times, which is exactly what one would predict if this conditioning reflected an invariant principle. Van Gelderen also provides evidence from the catastrophic loss of morphological case in English for distinguishing between universals of this type and the resetting of parameters, which appears to be faster and not predictable in historical corpora – no matter what style of language they reflect. As regards the latter, Van Gelderen's account is important since it demonstrates unequivocally the significance of choosing the right kind of data for documenting syntactic change in real time (cf. Bauer 2002 and Van der Wurff 2000). Despite its being derived from spoken data, the formality of the *CSE* by comparison to the *BNC*, as illustrated in (2a/b) below, appears to be so unduly influenced by prescriptive case rules that it does not make a good comparator for assessing the trajectory of syntactic change affecting nouns and pronouns. For different reasons, the same could be said of Shakespeare's *First Folio* when compared to the *HC,* since the former is dramatic verse, which may or may not be a true reflection of actual speech habits in the Early Modern period (see Hope 1994 and Kytö & Walker 2003).

(2) a. The change of pace to which *Barbara and I* are looking forward with
         real relish   (*CSE*-FAC95)
     b. *Me and my mother* have erm arranged it all   (*BNC*-KC8 920)

As noted briefly earlier, stylistic differentiation is also a concern of Cheshire's contribution which examines *inter alia* the preponderance of prefabricated expressions in spoken as opposed to written English. Her particular focus is: (1) the variation between conventional subordinating adverbial *when* clauses and lone *when* clauses; (2) the pronoun tags often found in Northern British Englishes and (3) verbal -*s*. All three are typical of spoken utterances and occur in a range of contemporary English corpora including her Reading corpus (see Cheshire 1982) and the Hull database collected in the 1990's (see Cheshire, Kerswill, & Williams 1999). In addition to her narrow analysis of 'style' (mentioned earlier with respect to gender preferences for different kinds of narrative) she tackles the broader issue of spoken versus written data also raised in Van Gelderen's diachronic account and Henry's synchronic account. Cheshire's chapter forcefully demonstrates that certain constructions which occur only in spoken language have not been taken seriously enough by either variationists or generativists, yet such data may well be crucial in answering some of the questions posed in §§3.1–3.2.

Given the fact that contemporary sociolinguistics evolved from traditional dialectology (cf. Britain 2002 and Chambers & Trudgill 1998), we were keen to commission some chapters in the present volume that would focus on the spatial diffusion of linguistic variation while drawing on the formalisms of generative theory. Although a number of the papers contain detailed information regarding the geographical setting of their investigations (Adger & Smith, King and Muysken, for example) and there are others in which particular regions are isolated as being associated with certain variable features (Cheshire and Henry), the contributions by Barbiers and Benincà and Poletto provide the most geographically informed socio-syntactic analyses contained in the volume.

As noted in §2, Barbier's chapter focuses on word order strategies associated with three-verb clusters in Dutch dialects. Although there is not space here to even outline the various geographical patterns uncovered in this research, we would like to highlight two important aspects of this investigation from the perspective of external and internal constraints on variability. Firstly, Barbier's findings with respect to the ordering of verb clusters suggest that the categorical or variable use of the order in the verbal cluster is conditioned by morphosyntactic features, e.g. morphology of the verb and type of auxil-

iary condition. Hence, all of the vernacular speakers in question, irrespective of their spatial location, had exceptionally high agreement rates with respect to ungrammatical orders, indicating that internal linguistic conditioning is at work here. Secondly, some areas are more homogeneous than other dialect areas since in the former a large number of speakers only tolerate a single order, signifying that these dialects may be more subject to supralocal norms (as in the case of the 3–2–1 order in Friesland, for example). Thus, for speakers of the other dialects who accept more than one order, it is invariably the case that the standard Dutch order is usually also included by informants, signalling that they can resort to both standard and non-standard varieties (see §1.3).

The contribution by Benincà and Poletto also investigates geographical microvariation, though this time the focus is within the linguistic zone in which dialects of the Romance family are spoken. Although it is somewhat less concerned with the external causes of regional variation than Barbier's account is, they are similarly able to track dialect isolates such as those of the Lombard region which have a unique variant acting as a marker for all persons. Benincà and Poletto also uncover universal properties of varieties within this language family (implicational scales), finding, for example, that there is no dialect in the region which has a vocalic clitic variant for all persons that can be extended to third person singular but not to third plural or *vice versa.*

As this brief overview demonstrates, those of our authors who address the major forms of externally-induced dialect differentiation and interpret patterns of correlation with respect to these do so with the same fervour that they explore issues of formal syntactic theory.

4.   Conclusion

In this retrospective and prospective review, we have identified three key areas in which we believe that this volume will contribute to the maturation of a paradigm for the investigation of syntactic variation, *viz.* (1) Methodological innovation, (2) New theoretical applications and (3) Modularity.

As regards (1), the open-ended nature of syntax and the significance of attending to language style (not forgetting the importance of considering the subjectivity of native speaker judgements) lead us to suggest that data from a range of different sources should be mined before proposing analyses for particular syntactic features. In the same vein, there is strong evidence to support the importance of utilising both qualitative and quantitative methods in seeking to understand the dynamics of syntactic variation and change within

both sociolinguistics *and* biolinguistics. As far as (2) is concerned, we think that this volume supports Muysken's (1995: 2) argument that the 'classic' sociolinguistic approach to the notion of 'variable' forces a perspective on syntactic variation in which it always implies isolated, loose elements. Availing of the formal apparatus provided by the generative paradigm which necessitates a more holistic view of the grammar and takes a keener interest in the acquisition process permits a sociolinguistic account in which one has a more robust view of exactly which variants really are "alternate ways of saying 'the same' thing" (Labov 1972: 118) and demonstrates just how this variability might be learned. Modularity, the third of our critical themes, takes the notion of an integrated approach to syntactic variation even further. Not only does it encompass the syntax proper but it also highlights the importance of examining interface levels between various sub-components of the grammar which may well prove to be crucial loci for variability.

All that being said, the publication of this volume is an important achievement for the progress of linguistic theory more generally and we believe that it is an even more crucial milestone in the coming-of-age of 'Socio-Syntax' as a discipline in its own right.

Acknowledgements

Liliane Haegeman, Anders Holmberg, Jim Miller, Jim Milroy, Herman Moisl, John Nerbonne, Susan Pintzuk, Geoff Poole, Shana Poplack, Ian Roberts, Sali Tagliamonte, Tonjes Veenstra and Martha Young-Scholten.

Notes

1. This term was first used by Dorian (1994) to describe variability within the East Sutherland Gaelic community she was studying in which a "profusion of variant forms" was tolerated over an extended period of time (1994:633). As Wolfram (2002:778) notes it is distinctive and, therefore, should be viewed as "separate from stylistic, geographic" and other types of variability.

2. Early references to the gulf in question are contained in the collection of papers from a 1987 *CLS* Parasession on variation in linguistic theory subsequently edited by Beals et al. in 1994.

3. Accounts such as Henry (1995) and those in Barbiers et al. (2002) mark the beginnings of this approach.

4. Sociolinguistic models have, in fact, had considerable success in unraveling variation across historical, geographical and social space as well as along stylistic continua, though these issues have largely been ignored within the biolinguistic programme.

5. Corrigan (2000a, 2003) has also used similar arguments in her accounts of small clauses and infinitives in a northern dialect of Irish-English and D'Arcy (2004), which is based on Canadian English data, provides evidence that previous analyses of discourse *like* are problematic in this respect too since they wrongly assume that all strings containing the feature derive from the same syntactic structure.

6. This view is largely the result of genuinely problematic phenomena such as underreporting/overreporting by data subjects in response to ideologies of various kinds (cf. Milroy 1987:149–150).

7. As argued by King, this volume, by Adger during the ICLAVE Workshop discussion and in his joint paper in the present volume, cf. Adger and Smith.

8. While the various roles associated with the cognition, semiotic and syntactic modules are relatively perspicacious, the reader should note that the 'interaction' module incorporates the speaker-hearer relationships, and notions of social identity often analysed within the fields of discourse analysis and pragmatics.

9. The possibility that his model has some universal veracity seems clear from even a cursory evaluation of the recursive capacities of humans, which would naturally appear to reside in both their syntactic and cognitive domains.

10. Though in this case it is linked to Muysken's model of cognition, namely regularisation which in this case refers to the use of an invariant verb form instead of a number of specific inflected forms, when forming verbs.

11. Interestingly, these findings are congruent with those from investigations of bilingual first language acquisition research undertaken by Hulk and Müller (2000); Müller and Hulk (2001) and Hulk et al. (2003). Moreover, Cheshire also reports in her case study in this volume that variability with respect to pre-fabricated expressions appears to be governed by conditions of use in a similar manner.

12. This may well be the left periphery which is linked to the specification of the illocutionary force of the utterance (see Cheshire this volume), and ultimately to the speaker's pragmatically-motivated choice. For instance, to put a constituent in focus, or to topicalize it (see Sorace this volume). As such, further investigation of this domain may prove to be very valuable in our search for a better understanding of the locus of variation.

13. Readers might benefit from knowing that within the sociolinguistic model, stylistic differences are often seen to be "the locus of individual's internalization of broader social distributions of variation" Rickford and Eckert (2001:1) and it is interesting to note that the contributions by Cheshire and Van Gelderen both appear to offer support for this view.

References

Adger, David & Jennifer Smith (2003). "Variation and the Functional Lexicon: A Case Study in Auxiliary Deletion". Abstract for the 'Syntactic Variation' Workshop at the *Second International Conference on Language Variation and Change*, Queen Mary, University of London & University of York.

Auer, Peter "Europe's Sociolinguistic Unity, or: A Typology of European Dialect/Standard Constellation". Ms.: www.fips.igl.uni-freiburg.de/auer/ICLAVE-Leuven+6-2002.pdf. University of Freiburg.

Barbiers, Sjef, Leonie Cornips, & Suzanne van der Kleij (Eds.). (2002). *Syntactic Microvariation*. Electronic publication of the Meertens Institute, http://www.meertens. knaw.nl/projecten/sand/synmic/. Amsterdam: Meertens Institute.

Bauer, Laurie (2002). "Inferring Variation and Change from Public Corpora". In Chambers, Trudgill & Schilling-Estes (Eds.), 97–114.

Bayley, Robert & Vera Regan (Eds.). (2004). *The Acquisition of Sociolinguistic Competence*. Special Issue of *Journal of Sociolinguistics, 8*(3), 323–433.

Beals, Katherine, Janette Denton, Robert Knippen, Lynette Melnar, Hisami Suzuki, & Erica Zeinfeld (Eds.). (1994). *Papers from the 30th Regional Meeting of the Chicago Linguistic Society, Vol. 2: The Parasession on Variation in Linguistic Theory*. Chicago: Chicago Linguistics Society.

Belletti, Adriana (1988). "The Case of Unaccusatives". *Linguistic Inquiry, 19*, 1–34.

Bobaljik, Jonathan D. & Dianne Jonas (1996). "Subject Positions and the Roles of TP". *Linguistic Inquiry, 27*, 195–236.

Britain, David (2002). "Space and Spatial Diffusion". In Chambers, Trudgill & Schilling-Estes (Eds.), 603–637.

Brown, Keith (1991). "Double Modals in Hawick Scots". In Peter Trudgill & Jack K. Chambers (Eds.), *Dialects of English: Studies in Grammatical Variation* (pp. 74–103). London: Longman.

Cedergren, Henriette & David Sankoff (1974). "Variable Rules: Performance as a Statistical Reflection of Competence". *Language, 50*, 333–355.

Chambers, Jack K. (1992). "Dialect Acquisition". *Language, 68*, 673–705.

Chambers, Jack K. & Peter Trudgill (1998). *Dialectology* (2nd ed.). Cambridge: Cambridge University Press.

Chambers, Jack K., Peter Trudgill, & Natalie Schilling-Estes (Eds.). (2002). *The Handbook of Language Variation and Change*. Oxford: Blackwell.

Chambers, Jack K. (2003). *Sociolinguistic Theory* (2nd ed.). Oxford: Blackwell.

Cheshire, Jenny (1982). *Variation in an English Dialect: A Sociolinguistic Study*. Cambridge: Cambridge University Press.

Cheshire, Jenny (1987). "Syntactic Variation, the Linguistic Variable and Sociolinguistic Theory". *Linguistics, 25*, 257–282.

Cheshire, Jenny (1999). "Spoken Standard English". In Tony Bex & Richard J. Watts (Eds.), *Standard English: The Widening Debate* (pp. 129–148). London: Routledge.

Cheshire, Jenny, Paul Kerswill, & Ann Williams (1999). "The Role of Adolescents in Dialect Levelling". Final Report submitted to the Economic and Social Research Council, London, June 1999. Ref. R000236180.

Chomsky, Noam (1991). "Some Notes on Economy of Derivation and Representation". In Robert Freiden (Ed.), *Principles and Parameters in Comparative Grammar* (pp. 417–454). London: MIT Press.

Chomsky, Noam (1993). "A Minimalist Program for Linguistic Theory". In Kenneth Hale & Samuel Jay Keyser (Eds.), *The View from Building 20* (pp. 1–52). Cambridge, MA: MIT Press.

Chomsky, Noam (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam (1999). *Derivations by Phase.* Cambridge, MA: MIT Press.

Chomsky, Noam (2000). "Minimalist Inquiries". In Roger Martin, David Michaels, & Juan Uriagereka (Eds.), *Step By Step: Essays on Minimalist Syntax in Honor of Howard Lasnik* (pp. 89–156). Cambridge, MA: MIT Press.

Chomsky, Noam (2001a). *Beyond Explanatory Adequacy.* Cambridge, MA: MIT Press.

Chomsky, Noam (2001b). "Derivation By Phase". In Michael Kenstowicz (Ed.), *Ken Hale: A Life in Language* (pp. 1–52). Cambridge, MA: MIT Press.

Cornips, Leonie (1998). "Syntactic Variation, Parameters and their Social Distribution". *Language Variation and Change, 10*(1), 1–21.

Cornips, Leonie (in press). "Intermediate Syntactic Variants in a Dialect – Standard Speech Repertoire and Relative Acceptability". In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky, & Ralf Vogel (Eds.), *Gradedness*. Oxford: Oxford University Press.

Cornips, Leonie & Karen P. Corrigan (2005). "Convergence and Divergence in Grammar". In Peter Auer, Frans Hinskens, & Paul Kerswill (Eds.), *Dialect Change: Convergence and Divergence in European Languages* (pp. 96–134). Cambridge: Cambridge University Press.

Cornips, Leonie & Cecilia Poletto (2005). "On Standardising Syntactic Elicitation Techniques. PART I". *Lingua, 115*(7), 939–957.

Corrigan, Karen P. (1997). "The Syntax of South Armagh English in its Socio-Historical Perspective." Ph.D. dissertation, University College, Dublin, NUI.

Corrigan, Karen P. (2000a). "What Are 'Small Clauses' Doing in South Armagh English, Irish and Planter English?" In Hildegard L. C. Tristram (Ed.), *Celtic Englishes II* (pp. 75–96). Heidelberg: C. Winter.

Corrigan, Karen P. (2000b). "Syntax and Sociolinguistics: Reconfiguring Janus Bifrons". Paper presented at the *Meertens Research Institute*, Amsterdam, October 2000.

Corrigan, Karen P. (2000c). "*What Bees to Be Maun Be:* Aspects of Deontic and Epistemic Modality in a Northern Dialect of Irish-English". *English World-Wide, 21*(1), 25–62.

Corrigan, Karen P. (2003). "*For-To* Infinitives and Beyond: Interdisciplinary Approaches to Non-Finite Complementation in a Rural Celtic English". In Hildegard L. C. Tristram (Ed.), *Celtic Englishes III* (pp. 318–338). Heidelberg: C. Winter.

Corrigan, Karen P. (to appear). *Parametric Variation within a Socially Realistic Linguistics: Syntactic Variation and Change in South Armagh English*. Publications of the Philological Society. Oxford: Blackwell.

Coupland, Nicolas (2003). Sociolinguistic Authenticities. *Journal of Sociolinguistics, 7*(3), 417–431.

D'Arcy, Alex (2004). "New Perspectives on Discourse *Like*". Paper presented at *New Ways of Analyzing Variation 33*, Ann Arbor, October 2004.

Dorian, Nancy C. (1994). "Varieties of Variation in a Very Small Place: Social Homogeneity, Prestige Norms and Linguistic Variation". *Language, 70*, 631–696.

Eisikovits, Edina (1991). "Variation in Subject-Verb Agreement in Inner Sydney English". In Jenny Cheshire (Ed.), *English Around the World* (pp. 235–255). Cambridge: Cambridge University Press.

Fennell, Barbara & Ronald Butters (1996). "Historical and Contemporary Distribution of Double Modals in English". In Edgar Schneider (Ed.), *Focus on the USA* (pp. 265–288). Amsterdam: John Benjamins.

Foulkes, Paul, Gerry Docherty, & Dom Watt (1999). "Tracking the Emergence of Structured Variation." In *Leeds Working Papers in Linguistics and Phonetics* (pp. 1–25). Leeds: University of Leeds (see also: www.leeds.ac.uk/linguistics).

Gervain, Judith (2002). "Linguistic Methodology and Microvariation in Language". MA thesis, University of Szeged.

Halle, Morris & Alec Marantz (1993). "Distributed Morphology and the Pieces of Inflection". In Kenneth Hale & Samuel Jay Keyser (Eds.), *The View from Building 20* (pp. 111–176). Cambridge, MA: MIT Press.

Harris, John (1985). "The Polylectal Grammar Stops Here". *Centre for Language and Communication Studies*, Occasional Paper, 13. Dublin: CLCS, Trinity College Dublin.

Hazen, Kirk (1996). "Dialect Affinity and Subject-Verb Concord: The Appalachian-Outer Banks connection". *South Eastern Conference on Linguistics Review, 20*, 25–53.

Henry, Alison (1995). *Belfast English and Standard English: Dialect Variation and Parameter Setting*. Oxford: Oxford University Press.

Henry, Alison (2002). "Variation and Syntactic Theory". In Chambers, Trudgill & Schilling-Estes (Eds.), 267–282.

Henry, Alison (in press). "Nonstandard Dialects and Linguistic Data". To appear in *Lingua*.

Hope, Jonathan (1994). *The Authorship of Shakespeare's Plays*. Cambridge: Cambridge University Press.

Hudson, Richard (1995). "Syntax and Sociolinguistics". In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, & Theo Vennemann (Eds.), *Syntax: Ein Internationales Handbuch Zeitgenössischer Forschung* (pp. 1514–1528). Berlin & New York: Walter de Gruyter.

Hulk, Aafke C. J. & Natascha Müller (2000). "Bilingual First Language Acquisition at the Interface between Syntax and Pragmatics". *Bilingualism: Language and Cognition, 3*(3), 227–244.

Hulk, Aafke, Janneke Peet, & Leonie Cornips (2003). "Acquiring the Syntax of *Beaucoup at a Distance* as a Bilingual Child: An Experimental Study". In Ana Teresa Pérez-Leroux & Yves Roberge (Eds.), *Romance Linguistics: Theory and Acquisition* (= *Current Issues in Linguistic Theory, Vol. 244*) (pp. 299–316). Amsterdam & Philadelphia: John Benjamins.

King, Ruth (1989). "On the Social Meaning of Linguistic Variability in Language Death Situations". In Nancy C. Dorian (Ed.), *Investigating Obsolescence* (pp. 139–148). Cambridge: Cambridge University Press.

King, Ruth (1994). "Subject-Verb Agreement in Newfoundland French". *Language Variation and Change, 6*, 239–253.

King, Ruth & Terry Nadasdi (1996). "Sorting Out Morphosyntactic Variation in Acadian French: The Importance of the Linguistic Marketplace". In Jennifer Arnold, Renée Blake, Brad Davidson, Scott Schwenter, & Julie Solomon (Eds.), *Sociolinguistic Variation: Data, Theory and Method. Selected papers from NWAVE-23* (pp. 113–128). Stanford: Centre for the Study of Language and Information (CSLI).

Kroch, Anthony (1989). "Reflexes of Grammar in Patterns of Language Change". *Language Variation and Change, 1*, 199–244.

Kuhn, Thomas S. (1970). *The Structure of Scientific Revolutions* (2nd ed.). Chicago: University of Chicago Press.

Kytö, Merja & Terry Walker (2003). "The Linguistic Study of Early Modern English Speech-Related Texts: How 'Bad' Can 'Bad' Data Be?" *Journal of English Linguistics, 31*(3), 221–248.

Labov, William (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William (1989). "The Child as Linguistic Historian." *Language Variation and Change, 1*, 85–94.

Labov, William (1994). *Principles of Linguistic Change: Internal Factors.* Oxford: Blackwell.

Labov, William (1996). "When Intuitions Fail". In Lisa McNair, Kora Singer, Lise M. Dolbrin, & Michelle M. Aucon (Eds.), *Papers from the Parasession on Theory and Data in Linguistics. Chicago Linguistic Society, 32*, 77–106.

Lavandera, Beatrice (1978). "Where Does the Sociolinguistic Variable Stop?" *Language in Society, 7*, 171–182.

Masterman, Margaret (1974). "The Nature of a Paradigm." In Imre Lakatos & Alan Musgrave (Eds.), *Criticism and the Growth of Knowledge. Proceedings of the International Colloquium in the Philosophy of Science, London, 1965*, Vol. IV (pp. 56–76). Cambridge: Cambridge University Press.

McDonald, Christine & Joan Beal (1987). "Modal Verbs in Tyneside English". *Journal of the Atlantic Provinces Linguistic Association, 9*, 43–55.

Meechan, Marjory & Michele Foley (1994). "On Resolving Disagreement: Linguistic Theory and Variation – *There's Bridges*". *Language Variation and Change, 6*, 83–85.

Meyerhoff, Miriam (2000). "The Emergence of Creole Subject-Verb Agreement and the Licensing of Null Subjects". *Language Variation and Change, 12*, 20–230.

Milroy, Lesley (1987). *Observing and Analysing Natural Language.* Oxford: Blackwell.

Montgomery, Michael & Stephen J. Nagle (1993). "Multiple Modals in Scotland and the Southern United States: Trans-Atlantic Inheritance or Independent Development?" *Folia Linguistica Historica, XIV*, 91–107.

Mufwene, Salikoko (1994). "Double Modals in American Southern English: How Peculiar are They?". *Contemporary Linguistics* (Working Papers of the Linguistics Department, University of Chicago), *1*, 89–109.

Müller, Natascha & Aafke Hulk (2001). "Cross-Linguistic Influence in Bilingual First Language Acquisition: Italian and French as Recipient Languages". *Bilingualism: Language and Cognition, 4*(1), 1–21.

Muysken, Pieter (1995). "Ottersum Revisited: Style Shifting and Code-Switching". Paper presented at the Dutch *Tweede Sociolinguïstische Conferentie*, Lunteren, The Netherlands, May 18–19, 1995.

Pesetsky, David & Esther Torrego (2004). *The Syntax of Valuation and the Interpretability of Features.* MA, MIT and U. MA, Boston.

Pintzuk, Susan (1995). "Variation and Change in Old English Clause Structure". *Language Variation and Change, 7*, 229–260.

Pollock, Jean-Yves (1989). "Verb Movement, Universal Grammar, and the Structure of IP." *Linguistic Inquiry, 20*, 365–424.

Poplack, Shana & Sali Tagliamonte (1989). "There's No Tense Like the Present". *Language Variation and Change, 1*, 47–84.

Rickford, John R. (1987). "The Haves and Have Nots: Sociolinguistic Surveys and the Assessment of Speaker Competence". *Language in Society, 16*(2), 149–177.

Rickford, John R. & Penelope Eckert (2001). "Introduction". In John R. Rickford & Penelope Eckert (Eds.), *Style and Sociolinguistic Variation* (pp. 1–20). Cambridge: Cambridge University Press.

Rizzi, Luigi (1997). "The Fine Structure of the Left Periphery". In Liliane Haegeman (Ed.), *Elements of Grammar: Handbook in Generative Syntax* (pp. 281–337). Kluwer: Dordrecht.

Roberts, Ian G. (1997). *Comparative Syntax.* London: Arnold.

Roberts, Julie (2002). "Child Language Variation". In Chambers, Trudgill, & Schilling-Estes (Eds.), 333–348.

Roberts, Julie & William Labov (1995). "Learning to Talk Philadelphian". *Language Variation and Change, 7*, 101–122.

Sankoff, David (1988a). "Sociolinguistics and Syntactic Variation". In Fritz Newmeyer (Ed.), *Linguistics: The Cambridge Survey, IV* (pp. 140–161). Cambridge: Cambridge University Press.

Sankoff, David (1988b). "Variable Rules". In Ulrich Ammon, Norbert Dittmar, & Klaus J. Mattheier (Eds.), *Sociolinguistics: An International Handbook of the Science of Language and Society* (pp. 984–997). Berlin: Mouton de Gruyter.

Sankoff, Gillian (1980). "Above and Beyond Phonology in Variable Rules." In Gillian Sankoff (Ed.), *The Social Life of Language* (pp. 81–93). Philadelphia: University of Pennsylvania Press.

Santorini, Beatrice (1993). "The Rate of Phrase Structure Change in the History of Yiddish". *Language Variation and Change, 5*, 257–283.

Schütze, Carson T. (1996). *The Empirical Base of Linguistics. Grammaticality Judgements and Linguistic Methodology*. Chicago: University of Chicago Press.

Sells, Peter, John R. Rickford, & Thomas Wasow (1996a). "An Optimality Theoretic Approach to Variation in Negative Inversion in AAVE". *Natural Language and Linguistic Theory, 14*, 591–627.

Sells, Peter, John R. Rickford, & Thomas Wasow (1996b). "Variation in Negative Inversion in AAVE: An Optimality Theoretic Approach". In Jennifer Arnold, Renée Blake, Brad Davidson, Scott Schwenter, & Julie Solomon (Eds.), *Sociolinguistic Variation: Data, Theory and Analysis* (pp. 161–176). Stanford: Centre for the Study of Language and Information (CSLI).

Smith, Neil (1989). *The Twitter Machine: Reflections on Language*. Oxford: Basil Blackwell.

Tagliamonte, Sali (1998). "*Was/Were* Variation Across the Generations: View from the City of York". *Language Variation and Change, 10*(2), 153–192.

Taylor, Ann (1994). "The Change from SOV to SVO in Ancient Greek". *Language Variation and Change, 6*(1), 1–137.

Trudgill, Peter (1974). *The Social Differentiation of English in Norwich.* London: Cambridge University Press.

Trudgill, Peter (1988). "Norwich Revisited: Recent Linguistic Changes in an Urban British Dialect". *English World-Wide, 9*, 33–49.

Van der Wurff, Wim (2000). "Variation and Change: Text Types and the Modelling of Syntactic Change". In Ricardo Bermúdez-Otero, David Denison, Richard M. Hogg, & Chris B. McCully (Eds.), *Generative Theory and Corpus Studies, A Dialogue from 10 ICEHL*, Topics in Linguistics 31 (pp. 261–282). Berlin & New York: Mouton de Gruyter.

Wilson, John & Alison Henry (1998). "Parameter Setting within a Socially Realistic Linguistics". *Language in Society, 27*, 1–21.

Winford, Donald (1996). "The Problem of Syntactic Variation". In J. Arnold, R. Blake, B. Davidson, S. Schwenter, & J. Solomon (Eds.), *Sociolinguistic Variation: Data, Theory and Analysis* (pp. 177–192). Stanford: CSLI.

Wolfram, Walt (2002). "Language Death and Dying". In Chambers, Trudgill, & Schilling-Estes (Eds.), 764–787.

Wolfram, Walt & Natalie Schilling-Estes (1995). "Moribund Dialects and the Language Endangerment Cannon: The Case of the Ocracoke Brogue." *Language, 71*, 696–721.

Aspects of modularity

A modular approach to sociolinguistic variation in syntax*

The gerund in Ecuadorian Spanish

Pieter Muysken
Radboud University Nijmegen

1. Introduction

This paper deals with the use of a complex construction, the Spanish gerund, in a highly stratified, partially bilingual (Quechua-Spanish), community in Andean Ecuador.[1] On the basis of the data presented, it is argued that a modular approach to syntactic variation is needed to explain the range of the variation encountered. First, a number of approaches to syntactic variation will be briefly reviewed, which are then evaluated in terms of three pairs of oppositions: choice versus determinism, derivational versus configurational paradigms, outer form manifestations versus inner form parameters. After a brief overview of the modular approach presented here, the speech community and the syntactic variable analyzed are presented. Then I turn to the quantitative results, which are then analyzed in a modular perspective.

2. Models for syntactic variation

Within generative linguistics (in a broad sense), a number of models have been proposed over the years to deal with syntactic variation. I will attempt to give a brief overview here of some of these models.

The first attempts involved rule ordering, as in Klima's (1964) treatment of the contrast between *the child whom you saw* and *the child who you saw*, in terms of ordering Case Marking before or after Wh-Movement. Presumably, in

the *whom* example, the question word has received accusative case before it is fronted, while in the other construction the reverse order of rules applies. This line was systematically taken up by Carden (1968, 1973), who argued that the different interpretations of *All the boys didn't leave* corresponded to different syntactic dialects, distinguished in terms of the ordering of rules such as Quantifier Lowering and Neg-Transportation. The syntactic dialects postulated by Carden were highly individual, and did not correspond to identifiable social units of any kind. Rule order played an important role in phonological change and variation as well (cf. Kiparsky 1968).

A slightly later development focused on optional rules (a more recent example of this approach is Meechan & Foley 1994), which soon developed into the variable rule paradigm. Here the variation lies in the application or non-application of a rule. Variable rules were used in Labov's treatment of different varieties of English with respect to negative attraction and negative concord (1972b). In the same vein, Naro and Lemle (1976) assumed an optional subject-verb agreement transformation in Brazilian Portuguese, subject to various quantitative constraints. A host of proposals followed, exploring the possibilities of variable rule methodology.

A special subclass of the optional rules in the variationist literature concerns local deletive rules such as *que* deletion in Montreal French (G. Sankoff 1973) and deletion of deictic elements in initial position in Dutch (Jansen 1978). In recent work, Henry (e.g. 2002) has continued to defend the possibility of accounting for variation in terms of optional rules.

A development which was foreboded in the work of David Lightfoot (1979) on modals in early modern English and subsequently taken up in work by Kroch (e.g. 1988; cf. also Taylor 1994) is the idea that grammatical variation is, in part, due to the existence of competing grammars in the speech community. In this view, variation is not a phenomenon within the grammar itself, but the result of bilingualism or bi-dialectalism within the speech community. At time period X, different people have different grammars, and from time period X onwards, one set of grammars replaces another.

An entirely different track has been taken in approaches referred to as 'lexical learning' (Borer 1984). Here syntactic variation is assumed to be reducible to lexical variation. Obvious earlier examples of this approach include the work on auxiliary selection in Montreal French (G. Sankoff & Thibault 1980), where the choice of *avoir* or *être* with different verbs is studied, but there is a rich tradition here, particularly when functional categories became the focus of scholarly attention, and variation in the use of functional categories was taken to be the source of much or all syntactic variation. A recent

quote from Sportiche (1995) illustrates this approach: "…. variation is essentially confined to the pairing between morphophonological properties and semantico-syntactic properties of morphemes."

The lexical perspective also finds support in other theoretical models. In an approach within the HPSG model, Bender (2000) argues that copula absence in AAVE is indeed a syntactic, rather than phonological, variable. She then proposes an analysis in terms of a phonologically empty verb or construction, and argues for a sign-based perspective on variation. The choice of a particular 'sign' or form can be viewed as a lexical choice, to be sure.

A logically separate but related view, very influential in the last twenty years of the previous century, is parameter theory, originally conceived of to deal with differences between languages (such as the Null-Subject Parameter; Rizzi 1982), but also successfully applied to account for variation within a language (cf. e.g. Cornips 1998). With the rise of functional categories as the primary trigger of syntactic operations, parameter theory has gradually been transformed into a theory about 'strong' and 'weak' properties of individual functional categories. These can be viewed as individual lexical features of elements, and hence parameter theory can be unified with lexical learning theory.

Another approach in this tradition, heralded, among others, by Goldberg (1995) is construction grammar, an approach which views much of our grammatical knowledge as knowledge of relatively fixed patterns, constructions, which are then combined to form actual sentences. From a construction grammar perspective, syntactic variation amounts to variation in repertoires of syntactic patterns, akin to lexical variation. Examples of this approach in the domain of syntactic variation include Verhagen (2000). Cheshire (1987) distinguished between pragmatic variation (communicative intent) and syntactic variation (equivalent constructions). Cheshire's earlier plea for more attention to the different options speakers have within their different repertoires, with accompanying constructions, lexical choices, speech rhythms, etc. could best be accommodated in the framework of a model in which styles and registers are defined as repositories for usage patterns, constructions, etc. To some extent, the work on syntactic output filters originally proposed in Chomsky and Lasnik (1977), now reformulated as PF-interface filters, also fits into this tradition.

The rapidly developing approaches to syntax within Optimality Theory lead to the analysis of variation in terms of stochastic and non-stochastic OT constraint orders (Bresnan & Deo 2001; Bresnan & Manning 2002). In the coming period, these approaches will start gaining in prominence, but their overall potential is hard to evaluate as yet.

This concludes my brief survey of approaches to syntactic variation. The question is how to evaluate them. While many descriptions have shown that there is extensive syntactic variation within speech communities linked to a variety of social characteristics of the speakers involved, it has been very difficult to link this finding to most current work in syntactic theory. There are three main obstacles, I think, making such a link difficult.

Choice versus determinism.   First of all, sociolinguistics, dealing as it does with 'alternative ways of saying the same thing' (Labov 1972a) implies a choice made by the speaker between various options (Van Hout 1984). Often, this is phrased in terms of the optional status of different rules, but the point is more general. Syntactic theory, at least in its current Minimalist form, departs from the notion of blind, deterministic application of a series of procedures given a certain starting point (Chomsky 1995). Thus a first obstacle to overcome is the apparent disparity between choice and determinism. One solution is placing the issue of choice outside the grammatical mechanisms employed, another one would be reintroducing choice into the grammar as such, through the mechanism of optional rules.

Derivational versus configurational paradigms.   In the late 1960s, when the study of variation came to the fore, both in syntax and in phonology, the model of generative grammar was derivational. There is an abstract representation (in phonology the underlying form, in syntax deep structure), which is semantically interpreted. This abstract representation is mapped onto a surface or concrete representation in a series of derivational steps (phonological rules and transformations). In the derivational paradigm the variable, Labov's 'the same thing', can be equated with the underlying lexical or structural representation, and variation ('alternative ways of saying ...') arises depending on whether a particular derivational rule is applied or not, or when two rules are reordered. Alternative surface representations are the variants. The main point of discussion at the time was whether the applicability of a particular rule should be quantitatively specified (Cedergren & Sankoff 1974) or left optional (Bailey 1973; Bickerton 1971) in the formal grammar. Thus, in the derivational paradigm, the formal model and the intuitive notion of variation were reconciled. However, in many later models, the idea of a semantically interpreted underlying representation with variant surface manifestations was abandoned in favour of a configurational model, in which a single representation is subject to various constraints. In non-derivational models the original intuition did not have a place any more.

Outer form manifestations versus inner form parameters. A third obstacle comes from the fact that the classical sociolinguistic variables that have been studied involve outer form, to borrow the expression from Von Humboldt (1848): highly visible (though not always consciously perceived) aspects of language, such as pronunciation, intonation, lexical choice. On the other hand, the type of variation studied within the generative paradigm involves inner form, complex parameter settings with consequences which are mostly unnoticeable. This has the effect that the syntactic variation encountered may be perhaps be regional, but generally does not correlate directly with specific social class lines or stylistic choices, and may even be entirely individual, with variants randomly distributed through the speech community. In other words, it is often variation without social meaning. Thus the third obstacle to overcome involves finding instances of variation which are frequent and with immediately perceivable effects, while at the same recognizing that these instances may reflect more deep-seated differences within the language.

I will try to argue that the analysis of the variable studied in this paper from a modular perspective allows us to overcome these obstacles.

3.   Modularity

The approach taken here is modular in that it assumes that the human linguistic capacities do not form a single monolithic whole, but fall into several groups or clusters (Fodor 1983; Anderson 1993). Some of the clusters relevant here are:

Syntax (Chomsky 1995). The syntactic capacities may be the only ones which are strictly linguistic, and do not play a role outside of language processing proper. Our syntactic capacities are responsible for the fact that most linguistic structures, in different domains, are endocentric or headed. Furthermore, they are probably responsible for the fact that in language we have effects of 'movement' (such as the placement of question words in initial position). It is a matter of debate whether the recursive capacities of humans also fall under this syntactic capacity, since they occur elsewhere as well.

Semiotics (Barthes 1967; Halliday 1978; Mayerthaler 1988). In addition to syntactic capacities, semiotic capacities play an important role in language processing, even though they are not unique to it but also play a role in other human sign systems. Characteristic properties of the semiotic dimension of language are the distinctiveness of linguistic forms, the transparency of many

linguistic structures, the elementarity of units which undergo operations as single wholes, and the process of analogy operant in the formation of many new structures.

Interaction (Goffman 1967; Sacks 1994; Schiffrin 1994). An important role in language processing is played by our capacities for interaction. Many human interactions are linguistic, but many are not. The interactive capacities are responsible for the management of the sequential nature of information exchange, and for the cohesion of human discourse.

Cognition (Clark 1996; Jackendoff 2002). Finally, there are the crucial cognitive capacities that form an important part of our linguistic capabilities. A number of properties of our general cognitive system play an important role within language as well: the embedding of one unit inside of another one, the possibility of recursion of structural processes, the opposition between concepts along the dimension of an often binary feature, the displacement of language from the here and now. All these properties are central to language, and derive from more general cognitive capabilities.

In the discussion of the variation encountered, I will return to these four components of human linguistic capacities.

4.   The speech community

The speech community I will illustrate my analysis with is a highly stratified rural town in the central highlands in Ecuador, called Salcedo.[2] At the time of study it had, including its surroundings, around 8.000 inhabitants. It is situated in the Interandean valley, at an altitude of 2.800 meters, and mostly functions as a market town. With its immediate surroundings it contains three ethnic groups: the socially dominant *blancos* or *mestizos,* local Indians or *runas* from nearby communities, and *runas* from further afield, working as *cargadores*. These ethnic groups are represented in three triangles or pyramids. The top of the highest pyramid links up with the national prestige identity symbolized by the capital, Quito, while the bottom of the two lowest pyramids is asociated with *runa*, highland Indian, identity. The two larger groups are further stratified socio-economically and in terms of education, yielding a total of six more or less clearly defined classes (Figure 1).

Salcedo is an asymmetrically bilingual town: many Indians speak Quechua, and all of them at least some Spanish. The non-Indians know limited amounts

The national capital

GEN   gente   local elite

COM   comerciantes   shopkeepers

ALB   albañiles   skilled workmen

CHO   cholos   urban unskilled

CAM   campesinos   peasants

CAR   cargadores   load bearers

GEN

COM

CHO

ALB

CAM

CAR
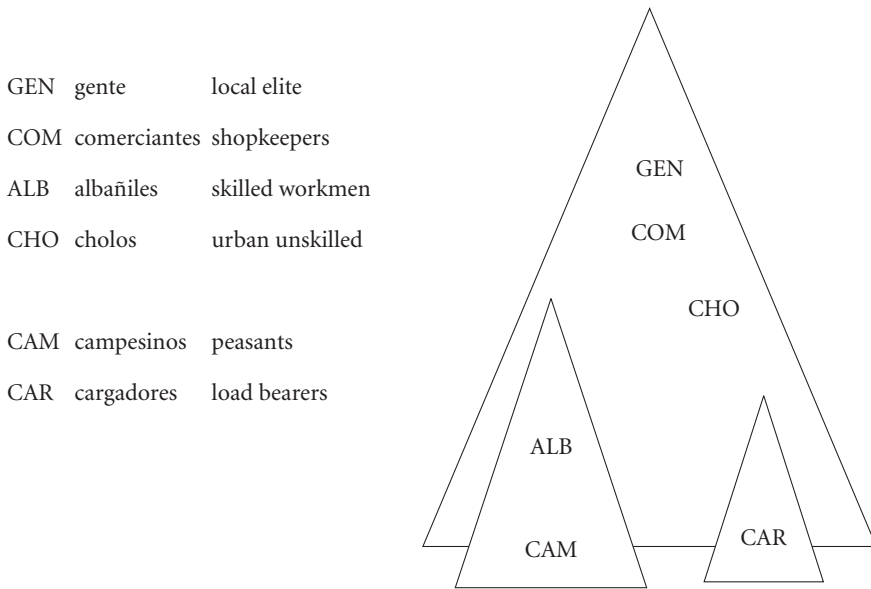
Indian "runa" identity

Figure 1. Schematic representation of the speech community

of Quechua, from rudimentary conversational knowledge to some words. In public life, at the time of study, only Spanish was heard. (The situation has changed somewhat in recent years due to the bilingual education movement, which has augmented the public stature of Quechua, if not its actual use.)

I will briefly describe the six classes one by one; the labels are my invention, but probably would be recognized by many. Of course, other divisions than the six-way split used here could be imagined, but for my purposes it will do. In any way, the ethnic split in three groups: 'whites', local Indians from nearby communities, migrant Indians from further away communities, is undisputable, with the proviso that the two types of Indians are both representatives of the *runa* identity, and that the 'whites' should not be perceived as necessarily racially Hispanic, although a few of them might be.

The *gente* or *gente decente* represent the local elite, with roots in the traditional land owning class. In addition to land holdings, they derive income from the higher administrative positions, often contained through connections in the capital. They live in the centre, often around the main square, the *parque*

'park'. Their wealth should not be exaggerated; their most valuable possession would be a family car. They generally will have had some secondary, in some cases even post-secondary, education. Members of this group with traditional land holdings may know some Quechua, but the others do not.

The *comerciantes* 'traders' may have a car, but this is a pick-up truck, used for transport of local produce and merchant goods, but also doubling as a collective taxi transportation vehicle for Indians from the nearby communities. They often own a store in town, work as school teachers, etc. They live near the centre, often nearby the market place. Their family connections are local or at most regional. This group will generally have had primary education, some secondary education, and in some cases also been to a teacher training college. This group may have some conversational knowledge of Quechua.

The *albañiles* 'construction workers' are Indians and have their house and family in the Indian communities rather than in town. However, they have generally worked outside of the region for longer periods, are skilled as construction workers, and receive reasonable salaries. This group has had at least some years of primary education, knows Quechua, but may not speak it from day to day.

The *cholos* 'lit. Indians with mestizo dress' live on the outside of the town, have few skills, and often marginal occupations. The term *cholo* is disrespectful, and could be glossed as 'half breed' if racial terms were not meaningless in a society where most people, even if so-called *mestizo* or *blanco,* are largely of Indian extraction. The feminine diminutive *cholita* is a term of endearment, however, and on the whole this group constitutes the main basis for Ecuadorian highland popular culture. In this group we find both people with and without primary education. There is some knowledge of Quechua, but as a whole this class is Spanish-speaking.

The *campesinos* 'peasants' live in the Indian community and derive most of their income from small plots of land, working on the large haciendas, and traditional artisan work. They tend to stick to traditional dress and modes of behaviour, also in their interaction with townspeople. Their main language is Quechua. The older *campesinos* are all illiterate, while younger ones may have had a few years of primary school.

The *cargadores* 'load bearers' come from largely monolingual Quechua-speaking Indian communities higher up on the mountains to work for months on end as load bearers, human donkeys, living in often appalling conditions. Literacy varies in this group.

As we will see, these six groups often differ considerably in their use of Spanish.

5.   The syntactic variable: Gerunds

The syntactic variable I want to consider is the use of the gerund form of the Spanish verb, which ends in *-ndo*. Like the English gerund, this form occurs in many constructions and is subject to much variation. The gerund is selected here as a variable both because of its grammatical complexity and because it is recognized by most Ecuadorians as a feature typical of bilingual highland Spanish, in some of its uses.

Before turning to the variation encountered, it is best to give an overview of the constructions involved. I will mark the examples that are also standard Latin American Spanish with an [S].

Some speakers allow gerunds on the main verb in independent clauses:

(1)   a.   *Tranquilo anda-**ndo***   [0 … V-ndo]
           Quiet   walk-GER
           "I walk quietly."

      b.   *Esos   roza-**ndo**, roza-**ndo***   [NP … V-ndo]
           DEM.M.PL clear-GER clear-GER
           "They keep clearing the grass."

In (1a) the subject is absent and needs to be recovered from the context, while in (1b) it is explicitly mentioned. In both sentences the tense needs to be inferred from the context; there may be a tendency to use gerunds in main clauses with continuous actions but this is not obligatorily the case.

A second, very productive, use of the gerund is with an auxiliary second verb. This is most common with the verb *estar* 'be', as in (2a), generally marking progressive. In (2b) the verb *ir* 'go' is used, combined with the gerund to indicate concomitant movement, while in (2c) the verb *venir* 'come' is used to mark that the action of the gerund took place before the action of coming. Another use of *venir* 'come' is illustrated in (2d), to mark an incremental progressive. The combination of the gerund with the verb *dar* 'give' in (2e) indicates a benefactive, while the rather fixed combination *mandar pegando* 'send beating' in (2f) refers to the result of both verbs. Finally, there are some highly irregular uses of the gerund, e.g. as the complement of the verb *querer* 'want' in (2g), where ordinarily an infinitive would be used.

(2)   a.   *Solito   estaba   ye-**ndo** por   Tigua* [S]
           Alone.DIM be.PST.3SG go-GER through Tigua
           "All by himself he was walking through Tigua."

b. *Yo   mismo   voy   lleva-**ndo** de   aquí el   almuerzo*
1SG self   go.1SG take-GER from here the lunch
"I carry along lunch from here myself."

c. *De   ahí   vine   cobra-**ndo***
From there come.PSTD.1SG collect-GER
"I came from there after collecting (my pay)."

d. *Bloque todito   viene   desgrana-**ndo***
Block   all.DIM come.3SG crumble-GER
"The building blocks are all crumbling."

e. *Nos   da   traye-**ndo***
1PL.OB give.3SG bring-GER
"He does us the favour of bringing it."

f. *De   ahí,   nosotros manda-mos pega-**ndo***
From there 1PL   send.1PL   beat-GER
"From there, we send them away beating them."

g. *Algunos,   si quieren ye-**ndo** bañar,   saben   ir   a*
Some.M.PL, if want.3PL go-GER bathe.INF know.3PL go to
bañar
bathe.INF
"Some people, if they want to go bathing, are wont to go bathing."

$$\text{NP}_i \text{ V } [\text{ 0}_i \ldots \text{ V-ndo }]$$

Notice that in all cases the auxiliary verb precedes the gerund, and the understood subject of the gerund is identical to that of the auxiliary: NP_i V [0_i … V-ndo].

A third main usage of the gerund is in adverbial clauses, which can mark concomitant actions, causality, temporal sequence, conditionality, concessive, etc. (Stump 1985; Martínez 2000). These adverbial clauses differ along two dimensions: whether the adverbial gerund precedes the main verb or not, and whether the understood subject of the gerund verb is identical to that of the main verb or not. In example (3) the main verb precedes the gerund, and the subjects are identical [Subj_i Verb [0_i V-ndo]]:

(3) *Y   vino   de   los EEUU finjie-**ndo**   que era*
And come.PSTD.3SG from the US   pretend-GER that be.PST.3SG
una media gringa [S]
one half gringo.F
"And she came from the US pretending that she was half a gringa."

$$[\text{ Subj}_i \text{ Verb } [\text{ 0}_i \text{ V-ndo }]]$$

In the following case, (4), the subjects are identical, but the gerund precedes the main verb [[0$_i$ V-ndo] Subj$_i$ Verb]:

(4) *Entra-**ndo** del   puente así   se   va   al*
     Enter-GER from.ART.SG.M bridge thus RE go.3SG to.ART.SG.M
     centro [S]
     centre
     "Entering from the bridge thus one goes to the centre."
                                                     [[0$_i$ V-ndo] Subj$_i$ Verb]

In (5), which contains two gerunds, the main verb comes first, and has a subject distinct from that of the first gerund, which, in turn, has a subject distinct from that of the second gerund [Subj$_j$ Verb [0$_j$ V-ndo]]: the so-called absolute construction. The person eating is distinct from the person advising, who again is distinct from the person going:

(5) *Al   ansioso come conseja-**ndo** ye-**ndo***
     To.ART.SG.M greedy.M eat.3SG advise-GER   go-GER
     "(The bear)$_i$ ate the greedy person$_j$, who$_j$ had gone, when someone$_k$ had
     advised him$_j$."   [Subj$_i$ Verb [0$_j$ V-ndo]]

Thus it can become quite complicated to infer the identity of the different subjects, which can only be established in terms of the verb meanings and the discourse context.

   In (6) the subject of the gerund is distinct from that of the main verb, and the gerund precedes the main verb [[0$_j$ V-ndo] Subj$_i$ Verb]:

(6) *En cambio mata-**ndo** ya   te   ponen preso*
     In contrast kill-GER   already 2SG.OB put.3PL captive
     "In contrast if you kill someone, they put you in jail."
                                                     [[0$_j$ V-ndo] Subj$_i$ Verb]

In addition to the adverbial use of the gerund illustrated in (3)–(6), we find gerunds sometimes in reduced relative clauses, where they modify a noun phrase, as in (7):

(7) *una   chica vinie-**ndo** de   Latacunga* … [S]
     ART.IND.F girl   come-GER from Latacunga
     "A girl coming from Latacunga."   [NP$_i$ [0$_i$ … V-ndo]]

We also find cases such as (8), where the gerund occurs in the (often durative) complement of a perception verb:

(8) *Yo   hace   unos   años   le   vi   baila-**ndo** chumado*
     1SG makes   some   year.PL 3SG.OB see.PSTD.1SG dance-GER   drunk.M

     así [S]
     thus
     "I a few years ago saw him dancing, drunk like this."

                                                    V NP$_i$ [0$_i$ ... V-ndo]

Finally there are several cases where the gerund occurs as the complement of a preposition (8), where ordinarily an infinitive would be used, or in a non-finite question (10), where also ordinarily an infinitive would be in place:

(9) *Entonces desde comenza-**ndo** desde ya   no   como es   allá,*
     Then   since begin-GER   since already   not like be.3SG there

     veo   que   voy   a   ganar   más.
     see.1SG that go.1SG to earn.INF more
     "Then since beginning since it is not like that any more there, I see that I
     am going to make more."   [[P 0 ... V-ndo] ...]

(10) *No   sé   cómo mostra-**ndo** la   patrona*
     Not know.1SG   how   show   ART.F.SG mistress
     "I don't know how to show it to the mistress."
                                        [Subj$_i$ V ... [Wh/Conj 0$_i$ ... V-ndo]]

All the evidence presented here should make it clear that the gerund in the Spanish studied here as a wide variety of uses. Some of these correspond to standard Latin American Spanish, others are limited to Ecuadorian non-standard varieties, or even to non-native varieties. However, I will not discuss this issue in detail here, since it will become clear that even uses deviant from the perspective of the norm may have a wide distribution.

6. The distribution of the gerund and factors involved in the variation in use

The first striking thing about the use of the gerund is that it has a very different distribution in the various classes in the community studied. Table 1 presents the total number gerund forms in the data-set, expressed as a percentage of the total number of non-main verbs. It has a range from 12% with the *gente*, to 37% with the *cargadores*, the other groups lying in between. Note that the Indian *albañiles* have a lower percentage than the ethnically non-Indian *cholos*

Table 1. Percentage of gerund forms of the total of non-main verbs (gerunds, infinitives, finite subordinate

	Gente	Comerciantes	Albañiles	Cholos	Campesinos	Cargadores
% gerund	12	17	20	23	33	37
n =	1491	836	1279	398	448	440
n gerund =	176	149	254	94	148	167

Table 2. Frequency in absolute numbers of types of gerunds

	Gente	Comerciantes	Albañiles	Cholos	Campesinos	Cargadores
main	4	3	4	6	10	46
adverbial	50	45	78	30	44	68
verb comp	122	98	173	63	104	98
other	4	6	3	1	3	1

(20 versus 23). The overall difference in gerund use is reflected in a stereotype about gerunds as typical of 'Indian' Spanish.

However, Table 1 offers no specific information about the actual distribution of the different gerund constructions in Ecuadorian Spanish. Therefore, Table 2 provides information about the frequency of the types of gerund. It is striking that the main verb use of the gerund is only somewhat frequent with the *campesinos* (10 cases) and still more so with the *cargadores* (46 cases). Main verb gerund use may thus be considered a feature typical for Quechua dominant bilinguals, and then mostly for incipient bilinguals, i.e. the *cargadores*. In absolute figures, it is clear that the use of the gerund in adverbial constructions (or as the complement of an auxiliary verb), occurs with all classes with considerable frequency (at least when calculated in terms of absolute figures). The other uses, in relative and perception clauses, and in other complement structures, are not frequent enough in any group to be seriously studied any further from this perspective.

The next table, Table 3, provides the same information in terms of the percentage of non-main verbs of the different types of gerund. For main verb gerund use, the picture does not change, but it is clear that there is a steady progression in terms of the different classes of the use of adverbial gerunds, from 3% for the *gente* to 15.5% for the *cargadores*. The other groups fall neatly in between, and the *albañiles* again are slightly more standard in this respect than the *cholos*. The same observation holds for verb complement use: there is a steady progression from 8% to 22%; the only deviation here is that the *campesino* group scores one percentage point higher than the *cargadores*.

Table 3. Frequency of the types of gerunds in percentage of non-main verbs

	Gente	Comerciantes	Albañiles	Cholos	Campesinos	Cargadores
main					2	10.5
adverbia	3	5	6	7.5	10	15.5
verb comp	8	12	13.5	16	23	22
other						

All the analyses presented so far simply looked at gerunds, without taking into account the alternative to them. However, if we take the idea seriously that we should be looking at alternative ways of saying the same thing, a different type of analysis would be called for. For the use of the gerund as a main verb it is almost impossible to study the alternative for purely quantitative reasons: the use of main finite verbs would be quantitatively overwhelming. For the use of gerunds as the complement to auxiliary verbs it is also not at all straightforward to study the alternatives. The progressive construction with *estar* + gerund is the basic way of expressing progressive in all Spanish varieties, and the other constructions are individually not frequent enough to be studied systematically on their own across six different speaker groups.

This leaves us with the gerund in adverbial clauses, where it is possible to consider alternatives. However, even here the expression "the same thing" should be taken with a considerable pinch of salt, since the gerund is much less explicit than its finite counterpart (see also Stump 1985; and the general discussion in Cornips & Corrigan 2005). Depending on the context, (11) could have any of the meanings in (12) and more:

(11)   *Llovie-ndo   vengo*
     Rain-GER   come.1SG

(12)   a.   *Aunque   llueve,   vengo*
         Though   rain.3SG come.1SG
         "Although it rains, I come."

      b.   *Como   llueve,   vengo*
         Since   rain.3SG come.1SG
         "Since it rains, I come."

      c.   *Cuando   llueve,   vengo*
         When   rain.3SG come.1SG
         "When it rains, I come."

      d.   *Si   llueve,   vengo*
         If   rain.3SG come.1SG
         "If it rains, I come."

e. *Despues de que llueve,   vengo*
   After   of that rain.3sg come.1sg
   "After it rains, I come."

In spite of this, I compared the use of gerunds and finite verbs in adverbial clauses, since they are the closest semantic equivalents to the gerund. Table 4 presents the effect of different variables on the selection of a gerundial versus finite adverbial clause, using the variable rule analytic techniques developed by David Sankoff (1978). It turned out that age, gender, and schooling did not have a significant effect, while language dominance and class did. Quechua-dominant bilinguals were much more likely to use a gerund than Spanish-dominant bilinguals, with the Spanish monolinguals as an intermediate group. Likewise there was, not surprisingly, a strong class effect. The *cargadores* were much more likely to use a gerund than the *gente*, with the other groups ranged in between. In this analysis, the *albañiles* were more prone to use a gerund than the *cholos,* but the difference is not very large. On the grammatical side, there were strong linguistic constraints on the variation. Gerunds are much more likely when the adverbial clause follows the main verb than when it precedes it. Similarly, across the entire corpus, gerunds were more likely to appear when the subjects were identical than when they were not.

Now that the basic outlines of the variation are established, I can return to the question of modularity. What dimensions of language use are responsible for the variation encountered?

Table 4. Effect of different variables on the selection of a gerundial versus finite adverbial clause

input:	.245				
gender:	men	.494	women	.506	
age:	young	.381	middle	.538	old	.582
schooling:	none	.463	elementary	.601	secondary	.435
language dominance:	Q-domin.	.686	Sp-domin.	.309	monol.Sp	.506
class:	Gente	Comerciantes	Albañiles	Cholos	Campesinos	Cargadores
	.323	.336	.492	.434	.648	.752

precedes main clause:	.346	follows main clause	.654
subject distinct	.282	subject identical	.718

not significant: age, gender, education

7.   Modular perspectives

Looking at the variation encountered slightly more abstractly, we can view the data from several perspectives, linking these up to the four modules referred to in Section 2: cognition, syntax, interaction, and semiotics.

A first perspective was already alluded to in Section 5, language learning, which is linked to our module of cognition. While gerunds occur as main verbs in a few cases in all social classes, as in (1a) and (1b), this option is clearly over-exploited by the *cargadores*. This has nothing to do with their native language Quechua as such, but reflects the results of a general learning strategy, namely regularisation: to use an invariant verb form instead of a number of specific inflected forms, when forming verbs (Andersen 1990).

In the specific case of adverbial clauses there may be direct Quechua influence, leading to a substrate perspective. In Quechua there are two gerund-like forms, *-kpi* and *-sha*, which are differentiated in terms of whether the subject of the adverbial clause is identical (*-sha*), as in (14), or not (*-kpi*), as in (13), with that of the main clause. Furthermore, Quechua adverbial clauses tend to precede rather than follow the main verb. Below we will test the hypothesis that Quechua-dominant bilinguals do indeed use more adverbial clauses with distinct subjects.[3]

(13)   *Tamia-**kpi**   mana shamu-sha-chu*
       Rain-GER.DS not   come-1.FU-NEG
       "If it rains I won't come."

(14)   *Puri-**sha**   shamu-ni*
       Walk-GER.SS come-1SG
       "Walking I come."

The substrate relates to reliance on knowledge of the native language in producing the second language.

A second way of looking at gerund use is the perspective of discourse and interaction (cf. e.g. the work on cohesion by Halliday & Hasan 1976). Consider an example such as (15) from one of the *cargadores*, in which a whole series of actions is linked together through gerunds.

(15)   *Todo i-**ndo**   a Machachi yo   solo   queda-**ndo** cocina-**ndo***
       All   go-GER to Machachi 1SG alone stay-GER   cook-GER

> *comemos   no   más*
> eat.1PL   not   more
> "When all had gone to Machachi and I stayed by myself, I just cooked and
> we ate."

Producing and interpreting such sentences in ordinary discourse requires a whole set of strategies of marking discourse cohesion, which interacts with structural features of the utterances.

The possibility of forming such utterances specifically relies very much on the perspective of interpretative strategies, where each strategy is based on a condition that needs to be fulfilled for the subject of the gerund to be identified. I will discuss four such strategies:

a. The subject of the gerund clause is interpreted through co-reference to a more prominent noun phrase that precedes it. This is a very standard condition, which holds for:

– gerunds as complements of auxiliaries;
– adverbial clauses following the main verb with a subject identical to that of the main verb;
– gerunds in relative clauses;
– gerunds in perception complements.

b. The subject of the gerund clause is interpreted through co-reference to a more prominent noun phrase that follows. This condition holds for gerunds in adverbial clauses with a subject identical to that of the main verb preceding the main verb.

c. The subject of the gerund clause is interpreted from the context. This condition holds for:

– main clause gerunds without a lexical subject;
– gerunds in adverbial clauses where the subject is distinct from that of the main clause but not lexically specified.

d. The subject of the gerund clause is a lexical noun phrase. This condition holds for:

– main clause gerunds with a lexical subject;
– gerunds in adverbial clauses where the subject is distinct from that of the main clause and lexically specified.

Table 5. Frequency in absolute figures of interpretive strategies for gerunds

	Gente	Comerciantes	Albañiles	Cholos	Campesinos	Cargadores
a.	164	120	198	83	126	114
b.	4	2	13	7	13	27
c.	7	16	8	5	15	55
d.	5	9	15	4	6	6

Table 6. Frequency in percentage of total number of gerunds of interpretive strategies for gerunds

	Gente	Comerciantes	Albañiles	Cholos	Campesinos	Cargadores
a.	92	82	85	84	79	56
b.	2	1	5.5	7	8	13
c.	4	11	3.5	5	9	27
d.	3	6	6	4	4	3

Notice that these strategies both appeal to hierarchy, through the notion of prominence, and to sequencing. I will assume that this sequencing is part of the discourse structure, and that the prominence is based on human syntactic competence (and hence is part of the syntax module), which defines asymmetric tree-structures.

First, in Table 5, the frequency in absolute figures of the different interpretive strategies for gerunds across the different speaker groups is given. Notice that strategy a. is well represented for all groups, while strategy b. is found mostly with the Quechua speakers, including the *albañiles* (13 cases with Spanish monolinguals, 53 cases with Quechua-Spanish bilinguals). This confirms the idea that adverbial clauses preceding the main verb may be a feature related to Quechua substrate.

A slightly different picture is given in Table 6, where the frequency is represented in percentages of the total number of gerunds of the four interpretive strategies for gerunds. The table shows that strategy a. characterizes over 90% of the gerunds used by the *gente*, but only 56% of the gerunds used by the *cargadores*. The other groups are all around 80%. Percentage-wise, strategy b. is mostly characteristic of the *campesinos* and the *cargadores*, but shows a steady progression across the classes. Interpretive strategy c., where the identity of the subject of the gerund is established contextually, is highest for the *cargadores* (27%), but also surprisingly high for the *comerciantes* (11%). Finally, strategy d., gerunds with a lexical subject, is not particularly frequent for any group.

Thus cognition, interaction, and syntax all play a role in accounting for the variation in the use of the gerund in Ecuadorian Spanish.

The fourth module, semiotics, also plays a role, if we assume that it is the use of the gerund form as such, – which may be viewed as a linguistic sign as in the work of Bender (2000) –, which functions as a linguistic variable. Support for this may come from the overuse of the gerund in a variety of constructions: where an infinitive would be called for, illustrated in examples (2g), (9), and (10) above. Notice that frequent use of the gerund in one construction appears to be linked to similar frequent usage of the gerund in other constructions. While the actual way the gerund is used is governed by abstract strategies, its very presence is indexical of social class and ethnic membership.

Returning now to the obstacles listed at the end of Section 2 of this paper, the analysis given here of the gerund in Ecuadorian Spanish avoids the problems listed there:

Choice versus determinism. The choice of the gerund form is, as such, either a lexical or constructional choice, and triggers a certain syntactic structure rather than being dependent of a choice on a different level.

Derivational versus representational paradigms. The strategies of interpretation perspective is non-derivational; the variation is not governed by choices in the derivational history of the constructions involved.

Outer form manifestations versus inner form parameters. The gerund verb form functions indexically at the level of the outer form of language, just like a word or a pronunciation would, even where abstract principles underly its use in particular constructions.

The different interpretative strategies for gerund use determine which gerund constructions are used by whom, but these strategies are established through the interaction between different modules of our linguistic competence. Currently we lack the appropriate formalisms to represent these interactions, but it may well be possible to devise a stochastic OT-model in which different rankings of the interpretative strategies as constraints reflect different grammars in the community.

Notes

1. Data on the Quechua spoken around Salcedo are presented in Muysken (1977). Other analyses of aspects of the Spanish of the area are presented in Muysken (1982, 1987), Van der Ent (2001) and Olberz (2002, 2003). A mixed, intertwined or split language spoken near Salcedo, as an in-group language mostly of the *albañiles*, called Media Lengua, is described in Muysken (1981, 1996).

2. The data were recorded on cassette in informal interviews and group conversations, in which there were generally 2–3 speakers present in addition to the author, making the recordings. Since the researcher had been working and living in the small town previously for two years gathering data on Quechua and Media Lengua, it was not difficult to make contacts and record informal Spanish.

3. This possibility is supported by the fact that in Media Lengua the use of *-kpi* and *-sha* alternates with that of *-ndu*.

References

Andersen, Roger W. (1990). "La Adquisición de la Morfologia Verbal". *Lingüística, I*, 89–141.

Anderson, John R. (1993). *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum.

Bailey, C.-J. N. (1973). *Variation and Linguistic Theory*. Washington, DC: Center for Applied Linguistics.

Barthes, R. (1967). *Elements of Semiology*. Translated by A. Lavers & C. Smith. London: Jonathan Cape.

Bender, Emily M. (2000). Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence. Unpublished doctoral dissertation, Stanford University.

Bickerton, Derek (1971). "Inherent Variability and Variable Rules". *Foundations of Language, 7*, 457–492.

Borer, Hagit (1984). *Parametric Syntax: Case Studies in Semitic and Romance Languages*. Dordrecht: Foris.

Bresnan, Joan & Ashwini Deo (2001). "Grammatical Constraints on Variation: '*Be*' in the Survey of English Dialects and Stochastic Optimality Theory". Ms., Stanford University.

Bresnan, Joan & Chris Manning (2002). "Stochastic OT and Syntactic Variation". Ms., NWAVE Workshop.

Carden, Guy (196). *English Quantifiers*. The Computation Laboratory of Harvard University, Mathematical Linguistcs, and Automatic Translation. Report No. NSF-20 to the NSF. Cambridge, MA.

Carden, Guy (1973). "Disambiguation, Favoured Readings, and Variable Rules". In C.-J. Bailey & R. W. Shuy (Eds.), *New Ways of Analyzing Variation in English* (pp. 171–182). Washington, DC: Georgetown University Press.

Cedergren, Henrietta & David Sankoff (1974). "Variable Rules: Performance as a Statistical Reflection of Competence". *Language, 50*, 333–355.

Cheshire, Jenny (1987). "Syntactic Variation, the Linguistic Variable, and Sociolinguistic Theory". *Linguistics, 25*, 257–282.

Chomsky, Noam (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam & Howard Lasnik (1977). "Filters and Control". *Linguistic Inquiry, 8*, 425–504.

Clark, Herbert H. (1996). *Using Language*. Cambridge University Press.

Cornips, Leonie (1998). "Syntactic Variation, Parameters, and Social Distribution". *LVC, 10*, 1–21.

Cornips, Leonie & Karen Corrigan (2005). "Convergence and Divergence in Grammar". In P. Auer, F. Hinskens, & P. Kerswill (Eds.), *Dialect Change: Convergence and Divergence in European Languages* (pp. 96–134). Cambridge: Cambridge University Press.

Van der Ent, Cecile (2001). El Español Andino Ecuatoriano; Un Estudio Sobre el Uso de los Clíticos *en Salcedo, Ecuador*. Unpublished MA-thesis, Leiden University.

Fodor, Jerry A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Goffman, Erving (1967). *Interaction Ritual*. New York: Doubleday.

Goldberg, Adele E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure* (= Cognitive Theory of Language and Culture Series). Chicago: University of Chicago Press.

Halliday, Michael A. K. (1978). *Language and Social Semiotic*. London: Edward Arnold.

Halliday, Michael A. K. & Ruqaiya Hasan (1976). *Cohesion in English*. London: Longman.

Henry, Alison (2002). "Variation and Syntactic Theory". In J. K. Chambers, Peter Trudgill, & Natalie Schelling-Estes (Eds.), *The Handbook of Variation and Change* (pp. 267–282). Oxford: Blackwell.

Van Hout, Roeland (1984). "The Need for a Theory of Choice in Sociolinguistics". *Linguistiche Berichte, 90*, 39–57.

Von Humboldt, Wilhelm (1836). *Über die Verscheidenheit des Menschlichen Sprachbaues und Ihren Einflusz auf die Geistlichen Entwicklung des Menschengeschlechts*. In Wilhelm von Humboldts gesammelte Werke, Vol. 6. Berlin: G. Reimer, 1848.

Jackendoff, Ray (2002). *Foundations of Language*. Oxford: Oxford University Press.

Jansen, Frank (1978). "Discourse as a Testing Ground for Syntactic Rules: Deletion of Sentence Initial Elements in Spoken Dutch". *Proceedings of the 12th International Congress of Linguists*, 544–548. Innsbruck.

Kiparsky, Paul (1968). "Linguistic Universals and Linguistic Change". In Emmon Bach & Robert T. Harms (Eds.), *Universals in Linguistic Theory* (pp. 170–202). New York: Holt.

Klima, Edward (1964). "Negation in English". In Jerold A. Fodor & J. Katz (Eds.), *The Structure of Language* (pp. 246–323). Englewood Cliffs, NJ: Prentice Hall.

Kroch, Anthony (1988). "Reflexes of Grammar in Patterns of Language Change". *LVC, 1*, 199–244.

Labov, William (1972a). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William (1972b). *Language in the Inner City*. Philadelphia: University of Pennsylvania Press.

Lightfoot, David (1979). *Principles of Diachronic Syntax*. Cambridge: Cambridge University Press.

Martínez, Glenn A. (2000). "A Sociohistorical Account of Grammatical Simplification: The Absolute Construction in Nineteenth Century *Tejano* Narrative Discourse". *LVC, 12*, 251–266.

Mayerthaler, Willy (1988 [1981]). *Naturalness in Morphology*. Ann Arbor: Karoma.

Meechan, Marjory & Michele Foley (1994). "On Resolving Disagreement: Linguistic Theory and Variation – *There's Bridges*". *LVC, 6*, 63–85.

Muysken, Pieter (1977). *Syntactic Developments in the Verb Phrase of Ecuadorian Quechua*. Dordrecht: Foris.

Muysken, Pieter (1981). "Half-way Between Quechua and Spanish: the Case for Relexification". In Arnold Highfield & Albert Valdman (Eds.), *Historicity and Variation in Creole Studies* (pp. 52–78). Ann Harbor: Karoma.

Muysken, Pieter (1982). "The Spanish that Quechua Speakers Learn". In Roger W. Andersen (Ed.), *Second Languages* (pp. 101–124). Rowley, MA: Newbury House.

Muysken, Pieter (1987). "Contactos Entre Castellano y Quechua en el Ecuador". In S. Moreno Yanez (Ed.), *Memorias del Primer Simposio Europeo Sobre Antropologia del Ecuador* (pp. 377–451). Quito: Ediciones Abya-Yala.

Muysken, Pieter (1996). "Media Lengua". In Sarah G. Thomason (Ed.), *Contact Languages. A Wider Perspective* (pp. 365–426). Amsterdam: Benjamins.

Naro, A. & M. Lemle (1976). "Syntactic Diffusion". *Papers from the Parasession on Diachronic Syntax*, 221–240. Chicago: Chicago Linguistic Society.

Olberz, Hella (2002). "Dar + Gerundio en el Español Andino Ecuatoriano: Sintaxis, Semántica y Origen". *Círculo de Lingüística Aplicada a la Comunicación, 12*.

Olberz, Hella (2003). "<Venir + Gerundio> en el Español Andino Ecuatoriano: Un Producto de Contacto Lingüístico". In Claus Dieter Pusch & Andreas Wesch (Eds.), *Verbalperiphrasen in den (Ibero-) Romanischen Sprachen / Perífrasis Verbals en les Llengües (Ibero-)Romàniques / Perífrasis Verbales en las Lenguas (Ibero-)Románicas* (pp. 89–103). Hamburg: Buske.

Rizzi, Luigi (1982). *Issues in Italian Syntax*. Dordrecht: Foris.

Sacks, Harvey (1994). *Lectures on Conversation*. Ed. by Gail Jefferson. Oxford: Blackwell.

Sankoff, David (Ed.). (1978). *Linguistic Variation: Models and Methods*. New York: Academic Press.

Sankoff, Gillian (1973). "Above and Beyond Phonology in Variable Rules". In C. J. Baley & R. W. Shuy (Eds.), *New Ways of Analyzing Variation in English* (pp. 44–61). Washington, DC: Georgetown University Press.

Sankoff, Gillian & Pierrette Thibault (1980). "The Alternation between the Auxiliaries *Avoir* and *Être* in Montreal French". In G. Sankoff (Ed.), *The Social Life of Language* (pp. 311–346). Philadelphia, PA: The University of Pennsylvania Press.

Schiffrin, Deborah (1994). *Approaches to Discourse*. Oxford: Blackwell.

Sportiche, Dominique (1995). "Sketch of a Reductionist Approach to Syntactic Variation and Dependencies". In Hector Campos & Paula Kempchinsky (Eds.), *Evolution and Revolution in Linguistic Theory. Essays in Honor of Carlos Otero* (pp. 356–398). Washington, DC: Georgetown University Press.

Stump, Gregory (1985). *The Semantic Variability of Absolute Constructions*. Dordrecht: Reidel.

Taylor, Ann (1994). "The Change from SOV to SVO in Ancient Greek". *LVC, 6*, 1–37.

Verhagen, Arie (2000). "Interpreting Usage: Construing the History of Dutch Causal Verbs". In Michael Barlow & Suzanne Kemmer (Eds.), *Usage-Based Models of Language* (pp. 261–286). Stanford, CA: CSLI Publications.

CHAPTER 3

Selective optionality in language development

Antonella Sorace
University of Edinburgh

1. Introduction

Experimental research on native speakers points to a distinction between violations of 'soft' constraints, which trigger gradient linguistic judgments, and violations of 'hard' constraints, which lead to categorical linguistic judgments (Sorace to appear; Sorace & Keller 2005). Hard constraints are purely structural in nature, while soft constraints tend to be associated with the mapping between syntax and lexical semantics, pragmatics, and information structure. It is argued in this paper that this distinction allows a more fine-grained analysis of syntactic variation in individual speakers than would it possible within current syntactic theories.

The particular perspective adopted here is developmental, and the focus is on adult second language (L2) acquisition and native language (L1) attrition. Variation in the unstable competence is widespread, but systematic and constrained. Drawing on data from near-native L2 speakers, it is shown that purely syntactic features are unproblematic at this stage but features at the interface of syntax and discourse (which have clear interpretive effects) present residual optionality. Data from speakers who have been exposed to a second language for many years display 'emerging' optionality in the native grammar with respect to precisely the same interpretive interface features, whereas structural constraints do not change as a result of attrition.

These data point to a remarkable convergence among data from different domains. In all these cases, it is the same features that present instability and variation: constructions that belong to the syntax proper are fully acquired in L2 acquisition and are retained in L1 attrition. In contrast, constructions that

require the integration of syntactic knowledge with knowledge from other domains present residual optionality in L2 and exhibit emergent optionality in L1 attrition. While no comprehensive explanation of these facts is currently available, the paper discusses the differences between 'representational' accounts, which put interfaces within the syntactic competence, and 'computational' accounts, which place optionality at the level of processing abilities.

The structure of this paper is as follows. First, the main findings of research on L2 endstate grammars are reviewed. It is then shown that the characteristics of residual L2 optionality are also found in L1 attrition: in both cases, optionality occurs at the interface between the syntax and other cognitive systems.

It is argued that narrowing down the domain of residual optionality to interface domains is a step forward, but raises the problem of what exactly an 'interface' is: specifically, the question is whether crosslinguistic effects leading to optionality occur at the level of L2 knowledge representations, or at the level of L2 computational/processing abilities. Both types of explanations have been proposed, and are reviewed here in turn. The final section explores the relevance of recent experimental research in psycholinguistics and cognitive neuroscience, and suggests that the issue of 'reduced input' in bilingual development may also play an important role.

2. Adult bilingualism: Ultimate attainment and effects on the native language

Recent research on post-pubertal adult second language (L2) acquisition has led to two important generalizations. First, the developmental paths and outcomes of L2 acquisition present similarities with those characterizing other types of language development. Second, the final outcome of L2 acquisition may be divergent from the outcome of first language (L1) development, but divergence is selective (i.e. is not found across the board) and appears to be constrained by universal principles.[1] These two generalizations are related; in fact, they are two sides of the same coin.

There has been a growing awareness among L2 acquisition researchers that a better understanding of the mechanisms and processes involved in L2 acquisition is more likely to be obtained if L2 acquisition is put in the wider context of research on language development. As Meisel (2001) puts it,

> There can be no doubt that there exist important commonalities shared in part or totally by the various types of language development. Consequently,

> the division of labor among research disciplines may represent an obstacle to achieving an adequate understanding of the driving forces and the mechanisms determining linguistic development.   (Meisel 2001: 1)

This paper presents evidence that there are indeed commonalities between L2 acquisition and other types of language development, such as non-pathological attrition in the L1 of speakers who have had prolonged contact with a second language. Perhaps paradoxically, these commonalities are most evident if one considers the patterns of differences between the most advanced endstate grammar attained by L2 learners and the corresponding monolingual L1 grammar. Hawkins (2001) has recently stressed the importance of understanding and explaining L1-L2 differences in terms of "...changes in the way Universal Grammar (UG) interacts with other components of the mind, or in terms of changes that occur in components of UG" (2001: 364).[2] He suggests that "poverty of the stimulus" studies, which focus on the similarities between L1 and L2 acquisition, have played a crucial role in establishing that UG constraints are available in L2 acquisition, and in countering claims that L2 acquisition is 'fundamentally different' from L1 acquisition: these studies have obtained plenty of evidence that L2 grammars are shaped by the same kind of constraints as all natural grammars (see White 2003 for a review).[3] However, only a focus on L1-L2 difference can help us understand how age impacts on acquisition processes and outcomes, in short, how the language faculty changes over the lifespan.

The differences examined in detail here are those found in the endstate grammar of very advanced or 'near-native' L2 speakers of Italian whose L1 is English. These features are compared with the L1 of native Italian speakers under attrition from long-term exposure to English. The comparison reveals that it is the same areas of grammar that appear to be *unstable* in language development and change, regardless of the circumstances in which development takes place. With respect to many of the phenomena described here, it is the same domains of grammar that show permeability to crosslinguistic influence. It is perhaps unsurprising that these same domains have independently been identified as more likely to exhibit gradience and variation in fully developed languages (Sorace to appear; Sorace & Keller 2005). If confirmed by further research, the discovery of converging findings from different areas of inquiry may be an important step towards a unified theory of language development. Meanwhile, there are several unresolved questions that have arisen from the study of patterns of differences in endstate grammars: evidence from psycholinguistics and the cognitive neuroscience of bilingualism (Felser et al.

2003; Hahne & Friederici 2001, among others) provides some indications of how these questions may be addressed.

2.1   End-state grammars

It is a fact that learning a language in adulthood normally leads to different outcomes from learning a language in early childhood. While the steady state of L1 acquisition is known in advance and is deterministically attained, the steady state of L2 acquisition seems to be extremely variable. However, there are people who start acquiring a second language in late adolescence and adulthood and become virtually indistinguishable from native speakers, at least in some respects (see discussion below). Do these very able learners attain the same type of mental representations of grammatical knowledge that monolingual speakers attain? Do they use the same cognitive resources? And do they apply the same processing strategies in using their knowledge? Addressing these questions requires studying learners who have stabilized at the highest competence levels.

Research specifically focused on ultimate attainment can tell us what kind of 'steady state' can be reached in non-primary language acquisition, whether such a state is quantitatively/qualitatively different from the monolingual steady state, and whether it obeys universal constraints. The characteristics of the best attainable final state – that is, the competence of near-native speakers – are, in a sense, more revealing of UG constraints on L2 acquisition than those of other stages (Borer 1996; Sorace 2000a, 2000b, 2003). If it is found, for example, that intermediate grammars appear to violate UG, the argument can always be made that, given more input, or more time, or a better learning environment, the non-native grammar may in due course converge on the target. However, adult learners who have reached the near-native level, and continue to benefit from full exposure to the L2, can be assumed to have progressed to the furthest attainable competence level: if there are differences between their grammar and the target grammar, these differences may therefore be considered as permanent. Investigating these differences is tantamount to testing the limits of L2 acquisition.[4]

Of the three possible ultimate attainment scenarios – convergence, constrained divergence, and unconstrained divergence (Sorace 1993; Sorace 2003; White 2003) – the focus here is a particular instantiation of constrained divergence: the scenario in which the endstate grammar exhibits *residual optionality* due to subtle and persistent L1 influence. This phenomenon will be illustrated with data from on-going research on pronominal forms (Filiaci 2003; Sorace

2000a, 2000b, 2003; Belletti, Sorace, & Bennati forthcoming) in null subject languages. Let us first examine how L2 speakers of Italian acquire the main feature of the 'null subject parameter' (Rizzi 1982), that is, the option of using both overt and null subjects.

As extensively argued in the theoretical syntax literature (Rizzi 1982; Belletti 2003), Italian is a 'null subject' language because it allows the omission of subjects in main clauses. The identification of null subjects is made possible by the presence of rich verbal morphology, which specifies person and number features. The distribution of null and overt subjects is regulated by discourse-pragmatic factors: subjects may be null when they refer to a topic, that is, an entity already introduced in the linguistic or situational context; overt subjects are used instead to introduce a new referent, or to contrast a referent with others. Thus, the learner of L2 Italian has to acquire both the syntactic parameter that licenses null subjects and the discourse-pragmatic conditions on subject realization. Both aspects may be presumed to be acquired at the near-native level, but in fact near-native speakers of Italian do not show identical behavior to native Italians.[5]

In response to a question such as (1a), English near-native speakers of Italian may optionally produce (1b), containing an overt subject pronoun, where a monolingual Italian speaker would have a clear preference for (1c), with a null pronoun.[6]

(1) a. Perchè Maria non ha parlato con nessuno?
        Why   Maria not has talked to   anyone?
     b. Perchè *lei* è troppo timida
        Because she is too   shy
     c. Perchè Ø è troppo timida
        Because Ø is too   shy

The reason why (1b) is anomalous is that there is a relationship of topic continuity with the previous sentence: Italian requires a null subject pronoun in this case (Grimshaw & Samek-Lodovici 1998, among others).

In contrast, the same near-native speakers never produce a null pronoun when there is a shift of topic, as in (2b), or when the subject is contrastive, as in (3b). Moreover, they do not introduce null pronouns in their English as a result of exposure to Italian, so (4b) is unattested.

(2) a. Perchè Maria non ha parlato con nessuno?
     b. *Ø (= Gianni) non l'ha neanche guardata
        Ø (= Gianni) didn't even look at her

(3) a. Maria ha detto che andava   da Paolo?
         Maria has said   that was going to Paolo's?
     b. *No, Ø (= Paolo) ha detto che   andava   da lei
         No, Ø   has said that (he) was going to her

(4) a. Why didn't Mary talk to anyone?
     b. *Ø was sick

A similar pattern is found with respect to another aspect of the null subject parameter: the relative position of the subject and the verb (see also Leonini & Belletti 2004). In response to an all-focus question, such as 'what happened', L1 English near-native speakers of Italian optionally place the subject in preverbal position regardless of the unaccusative or unergative status of the verb,[7] as in (5b)–(6b); native Italians, in contrast, would naturally place it after the verb, as in (5c)–(6c). This also happens in a contrastive context, such as (6), in which Italian requires the topic to be in postverbal position.

(5) a. Che cosa è successo?
         "What happened?"
     b. Gianni è arrivato
         Gianni is arrived
         "Gianni arrived."
     c. E' arrivato Gianni
         is arrived Gianni
         "Gianni arrived."

(6) a. Chi ha tossito?
         "Who coughed?"
     b. Gianni   ha tossito
         "Gianni has coughed."
     c. (Ha tossito)   Gianni
         "(has coughed) Gianni."

Furthermore, these speakers are more likely to produce sentences like (7a) regardless of whether the subject is definite or indefinite, whereas monolingual speakers would prefer a postverbal subject, as in (7b), particularly when the subject is indefinite (Belletti 1988).

(7) a. Hai sentito che un palazzo/il palazzo dell'ONU è crollato?
         "Have you heard that a building/the building of the UN collapsed?"
     b. Hai sentito che è crollato un palazzo/il palazzo dell'ONU?
         "Have you heard that is collapsed a building/the UN building?"

There are parallel effects in comprehension. For example, in the forward anaphora sentences in (8b), L2 near-native speakers of Italian are significantly more likely than monolingual Italians to judge the overt pronoun as coreferential with the matrix subject 'Maria', rather than with the complement 'la sua amica' or with a third extralinguistic referent; however, the null pronoun in (8a) is correctly interpreted as referring to the matrix subject.

(8) a. Mentre   attraversa la   strada, Maria saluta la sua amica
        While   Ø is crossing the street, Maria greets her   friend
     b. Mentre lei attraversa la   strada, Maria saluta la sua amica
        While   she is-crossing the street, Maria greets her   friend

A further distinction between preverbal and postverbal subjects is that preverbal indefinite subjects are interpreted as 'old' information (topic) whereas postverbal subjects are ambiguous between the two readings (Pinto 1997). This distinction is observed in the intuitions of native speakers of Italian, but near-native speakers have indeterminate intuitions and may therefore interpret the preverbal indefinite subject 'un bambino' in (9a) both as referring to one of the twins and as a new referent. In contrast, native and near-native speakers do not differ in their interpretation of (9b): both readings are allowed.

(9) Il mio vicino   del terzo   piano ha due gemelli
     My   neighbour on the third floor   has twins
     a. La   notte scorsa un bambino piangeva
        The night last   a   baby   was-crying
     b. La   scorsa notte piangeva   un bambino
        The last   night was-crying a   baby

A striking feature of these patterns is their asymmetry: near-native speakers of Italian overgeneralize overt subject pronouns and preverbal subjects to contexts which would require null subjects and postverbal subjects in native Italian, but they do not do the reverse, namely they do not extend null and postverbal subjects to inappropriate contexts. In fact, when they use null pronouns and postverbal subjects, they use them correctly. So these speakers do not lack *syntactic* knowledge: they have acquired a null subject grammar.[8] What seems to be at stake is knowledge of the appropriate felicity conditions for the use of overt subjects and preverbal subjects. The optionality in their grammar is at the level of the *discourse conditions* on the distribution of pronominals and on the placement of subjects. Thus, residual optionality primarily affects morpho syntactic features that are interpretable at the interface with conceptual systems (LF). The affected features may remain unspecified, giving rise

to optionality. Thus, residual optionality affects the use of overt subjects and preverbal subjects in L2 Italian, which is regulated by the interpretable [topic-shift] and [focus] features. If these features remain unspecified, overt subjects in near-native Italian are not necessarily being interpreted as shifted topics or foci.

3. Revisiting previous studies of near-nativeness

3.1 Interface divergences

A brief re-examination of the results previously obtained by studies on near-nativeness indicates a similar split between syntactic constraints and other types of interpretive conditions on the syntax.[9] Coppieters's (1986) classic study was the first investigation of competence differences between native and near-native speakers. He compared the judgments of native and near-native speakers of French on a variety of constructions, some of which he assumed exemplified 'formal UG properties' and others represented 'semantic' properties outside the domain of UG. While the overall profile of near native speakers was one of divergence from native speakers and lack of uniformity, it is intriguing that most differences were found with respect to the 'semantic' properties. For example, near-native speakers of French have significantly different intuitions from native speakers about certain interpretive contrasts between the imperfect and the present perfect. In (10), the present perfect imposes the meaning "Did you manage to drive in the snow?" whereas the imperfect in (11) means "Did you know how to drive in the snow?".

> (10) Est-que tu *as   su*   conduire dans la   neige?
>   is-that you have known-PERF drive   in   the snow

> (11) Est-que tu *savais*   conduire dans la   neige?
>   Is-that you knew-IMP drive   in   the snow

Native French speakers are consistently able to recognize this type of distinction, but near-native speakers have indeterminate intuitions, or actually attribute to one tense the meaning of the other. In contrast, the differences between the two groups with respect to purely syntactic properties (as for example what used to be called then the A-over-A constraint, in (12)–(13)) are much smaller.

(12) Cet homme, dont   j'admire le   tableau, est venu hier
      This man   of whom I admire the painting is   come yesterday

(13) *Cet homme, dont   je joue avec les enfants, est venu
      This man   of whom I play with the children is   come
      hier
      yesterday

Coppieters' study was criticized on methodological grounds (and was replicated in part by Birdsong 1992, who employed a more stringent methodology and did not obtain the same results).[10] However, it was probably right in pointing to a split between purely syntactic properties, for which L2 learners construct target-like representations, and properties responsible for interpretive differences, for which L2 learners construct divergent or indeterminate representations.

A more recent example of near-native divergence in the same domain is Montrul and Slabakova's (2003) study of tense/aspect interpretive distinctions in L1 English – L2 Spanish. Unlike Coppieters, Montrul and Slabakova consider that aspect falls within the domain of Universal Grammar, in line with recent theoretical research: specifically, it is determined by a functional aspectual category which is activated in Romance languages but not in English (Giorgi & Pianesi 1998). In Spanish, the preterit is used to mark perfective aspect, while the imperfect is used to indicate the unboundedness of the event. In English, on the other hand, past tense events are inherently perfective and unboundedness is expressed via the progressive. So the task faced by the English learner of Spanish is to acquire the [+/– perfective] distinction carried by the two tenses. In addition, they have to acquire the fact that the two tenses carry different, context-bound interpretations. The tense in Spanish impersonal constructions determines whether the null subject receives a specific or a generic interpretation. With the imperfect, both interpretations are available; with the preterit, only the specific interpretation is possible, as shown by the contrast in (14).

(14) a. Se comía   bien en ese   restaurante
          CL eat-IMP well in that restaurant
          "One/we ate well in that restaurant"
      b. Se comió   bien en ese   restaurante
          CL eat-PRET well in that restaurant
          "We/*one ate well in that restaurant"

This distinction is not taught, and tends to be acquired late. More importantly, the results of Montrul and Slabakova's study show that even though the major-

ity of near-native speakers of Spanish recognize it, and thus perform similarly to natives, a substantial minority (5 out of 12) do not. Hence, a minority of speakers at this advanced level exhibit a split between native-like syntactic representations of tense-aspect distinctions, and a divergent representation of interpretive properties related to these distinctions.

A pattern of residual optionality emerged from a study by Robertson and Sorace (1999) on V2 constructions in very advanced L1 German – L2 English. The overall outcome of this study shows that these learners, as a group, have acquired the fact that Standard English is a categorically non-V2 language, having divested itself of all but residual V2 effects since the Middle English period (Lightfoot 1999).[11] However, some of them occasionally produce sentences such as the following:

(15)   First of all one has to realize that in the past new developments always affected society. Whether it was the radio or the car it doesn't make any difference. *Always have been conservative warnings* that the harms would outweigh the positive consequences

(16)   Although in a highly developed country, like Germany, the majority of the people are well off, *for many kids is living with their parents* a nightmare

The explanation offered by Robertson and Sorace was in terms of a strong C feature (which is part of the interlanguage lexicon) wrongly entering the numeration for an English sentence: so, in a sense, the locus of optionality was assumed to be the interface between the syntax and the lexicon. But a different range of considerations may be made. German has generalized verb-second, in the sense that the verb always occupies the second position irrespective of the nature of the elements placed in clause-initial position. English, in contrast, exhibits residual verb-second, occurring only when certain types of constituents (usually negative adverbials and delimiting PPs) are fronted. This type of fronting is stylistically marked and relatively infrequent in the input. V2 phenomena are related to the specification of the illocutionary force of the utterance, and ultimately to the speaker's pragmatically motivated choice, for example, the decision to put a constituent in focus, or to topicalize it. However, it is only in English that these choices are lexically conditioned, producing a marked word order in a minority of cases. Full control of these conditions, and of their syntactic effects, is still problematic for a minority of German learners of English: optionality here involves the effects of the more consistent, generalised V2 system of German onto the less consistent, residual V2 constructions of English. Once again, we note a split between the native-like acquisition of

syntactic properties (i.e. a non-V2 grammar) and the residual optionality in an area that involves the complex interplay of syntactic and discourse conditions.

Hopp's (2004) study of Japanese and English near-native speakers of L2 German points to a discrepancy between knowledge of syntactic options and universal constraints on scrambling (a process that changes the order of constituents in a clause), which are completely acquired, and knowledge of the discourse-pragmatic conditions on the acceptability of scrambled orders, which remain indeterminate or divergent. For example, advanced Japanese speakers of L2 German acquire the strong uninterpretable syntactic feature that makes scrambling possible in the L2 (Oka 1996, among others), but their judgments diverge from those of native German speakers with respect to certain types of scrambling that are governed by extrasyntactic features. Thus, their acceptance of intact scrambling of indefinite DP over a definite DP, shown in (17), is much higher than for native German speakers, who assign a marginal acceptability status to this construction.

(17) Ich denke, dass einen Film über Frankreich Martin gestern
      I   think that a   film about France   Martin yesterday
      gesehen hat
      seen   has

According to Hopp, the source of the problem is a lower sensitivity to a contrastive focus constraint on this type of scrambling.

3.2 Interface convergences

In contrast with the research we have reviewed so far, there are some studies of near-native speakers that show no optionality or, in other words, complete convergence of near-native and native performance. One such study is Sorace's (1993) investigation of clitic-climbing and auxiliary change in L1 French – L2 Italian. In a verb complex consisting of a modal verb and an embedded infinitive, a clitic pronoun can be attached to the embedded verb, as in (18a), or it can 'climb' to a position preceding the main verb, as in (18b): if the embedded verb normally requires the perfective auxiliary *essere* 'be', there is an apparent optional 'transmission' of the auxiliary 'be' to the main verb (which normally selects *avere* 'have') when the clitic remains attached to the main verb: so both auxiliaries are possible, as in (18a). If there is clitic climbing, however, the auxiliary of the main verb obligatorily changes to 'be', as shown in (18b, c).

(18) a. Maria ha dovuto / è dovuta andarci
       Maria has had /   is had   go-there-CL
       "Maria had to go there"

  b. Maria ci   è dovuta andare
      Maria there-CL is had   go
      "Maria had to go there"

  c. *Maria ci   ha dovuto andare
      Maria there-CL has had   go

These phenomena, known in the literature under the name of 'restructuring' (Rizzi 1982; Burzio 1986), are not related to any interpretive differences or discourse conditions. The auxiliaries in (18a) often vary within individual idiolects; when there is a preference, this is not related to particular contexts.[12] Sorace (1993) found that French near-native speakers of Italian have a nativelike representation of clitic climbing (even though modern French does not allow this option) and acquire the obligatory auxiliary change triggered by this construction. The only difference between their intuitions and those of native Italians is that they have a strong preference for *avere* in (18a): in other words, they tend to eliminate the optionality of auxiliary choice.

Another study that found no differences between natives and near-natives (and no optionality) is White and Genesee (1996) on subjacency in L2 English. These authors investigated the competence of near-native speakers of English (selected via a complex screening procedure) with respect to different types of *wh*-extractions. The aim was to ascertain whether these speakers had acquired the subjacency constraints on *wh*-movement. Subjects performed in an acceptability judgment test that included stimuli like these:

(19) *What did you hear the announcement that Ann had received?

(20) *Who does Tom love the woman who married?

(21) Which car did the police claim Ann had stolen?

(22) Who did Jane announce would be the new teacher?

The constraints at work here are purely syntactic and do not interface with any lexical or pragmatic conditions. The results show that near-native speakers of English do not differ from native speakers in their judgments of complex extractions, and thus the constraints had been completely acquired.

4. Parallels between L2 acquisition and L1 attrition

There has been relatively little research on the effects of second language on the native language. The implicit assumption underlying much L2 acquisition research is that the native grammar usually does not change in response to L2 input. Much early research on attrition was concerned with migrant communities, which are usually characterised by diminished use of the L1, separation from the L1 speaking community, low degree of acculturation, and a low-level of L2 attainment at least in the first generation (see Weltens et al. 1986). The few studies on individual speakers (usually case studies), which have led to two important descriptive generalisations:

a. Attrition is a selective process: some aspects of the L1 grammars are more vulnerable than others. For example, Altenberg (1991) shows that case and number features in German are affected by English, but negative placement is not.
b. Attrition generally leads to the loss of restrictions on the application of rules. This generalisation was described in early attrition research. Seliger (1991) states that attrition involves the replacement of formally more complex and more narrowly distributed rules by formally less complex rules with wider distribution. In his 1991 study, Seliger found that the dative alternation rule in English L1, which is lexically governed, is affected by the more general and primarily syntactic L2 Hebrew rule.

Recent research within a generative grammar framework (Sorace 2000b; Tsimpli et al. 2003; Tsimpli et al. 2004) has explored attrition phenomena in more depth, focusing on the changes that occur in the pronominal system of native Italian speakers after prolonged exposure to English. The results of these studies indicate that native Italians who are near-native speakers of English exhibit an identical pattern of optionality as the English near-native speakers of Italian: these speakers may overgeneralize overt subjects and preverbal subjects to contexts which require a null subject or a postverbal subject, both in production and in comprehension. The reverse pattern is not found.

Thus, there is a parallelism between the end-state knowledge of English near-native speakers of Italian and the native knowledge of Italian near-native speakers of English under attrition with respect to null/overt subjects and pre/postverbal subjects. In both cases, the speakers' grammar is/remains a null-subject language. The computational features of syntax responsible for the licensing of null subjects are acquired completely: only the syntax-discourse interface conditions on pronominal subjects are affected by attrition.[13]

A similar discrepancy between syntactic properties and interface conditions is reported by Montrul (2004), who investigates the knowledge of pronominal forms (subjects, objects and clitics) in Spanish heritage speakers (i.e. second-generation Spanish-speaking immigrants to the United States). Her results indicate that purely syntactic properties are spared by attrition, whereas properties interfacing with discourse pragmatics and lexical-semantic factors are affected by it. For example, heritage speakers' knowledge of dative clitics in (23b) is identical to that of monolingual Spanish speakers, but their knowledge of the preposition *a* with direct objects in (24a, b) and clitic-doubling in (25), which are subject to factors such as animacy, lexical aspect, agentivity, and affectedness, is significantly more indeterminate.

(23) a. Patricia mandó una carta a mis amigos
          Patricia sent   a   card to my friends
      b. Patricia les   mandó una carta
          Patricia to them-DAT CL sent   a   card

(24) a. Inés conoce a varios aficionados
          Inés knows several fans
      b. La   opera conoce varios aficionados
          The opera knows several fans

(25) Cecilia le   lavó   las manos a Victoria
       Cecilia to her-DAT CL washed   the hands to Victoria

Heritage speakers also exhibit optionality with respect to pronominal subjects and subject-verb inversion, although not to the same extent as the Italian speakers in Tsimpli et al. (2003) (see also Gurel 2002 for attrition in Turkish subject pronouns and Polinsky 1995, in press, for attrition in Russian pronominal forms). In a different study of attrition in Spanish heritage speakers, Montrul (2002) reports attrition effects on the same tense-aspect distinctions illustrated in (14a, b) above, which create residual problems for individual L2 learners.

5. A generalization on optionality in bilinguals

At this point a generalization is needed that accounts for these results. As a first approximation, the following generalization may be proposed:

(26) "Narrow" vs. "Interface" syntax.

–   Features that are internal to the computational system of syntax proper are acquired successfully by adult L2 learners and are retained in the L1 under attrition; endstate (near-native) grammars converge with native grammars, and grammars under attrition do not diverge from monolingual grammars.

–   Features that belong to the interface between syntax and other domains, such as the lexicon, discourse, or pragmatics, may never be completely acquired by L2 learners and may be vulnerable to the effects of attrition. It is among these features that one finds 'residual' L2 optionality due to the influence of the native language and 'emerging' optionality due to the influence of the second language.

A closer look at the notion of 'interface' is in order. At first sight, it may appear as if the generalization in (26) contradicts decades of L2 acquisition research. Much early descriptive research on L2 acquisition, in fact, concluded that semantically more transparent properties are easier to learn than more abstract syntactic properties, which do not correspond in any clear way to semantic notions (see e.g. Kellerman 1987). Research on the 'basic variety' shows that early interlanguage grammars favor semantic and pragmatic principles of utterance organization (Klein & Perdue 1997). However, the generalization in (26) does NOT claim that syntactic aspects are easier than semantic aspects; rather, it suggests that aspects of grammar that require not only syntactic knowledge, but the ability to coordinate syntactic knowledge with knowledge from other domains is late acquired – in fact, possibly *never* completely acquired by L2 learners. The differentiation between syntactic computational properties and interface properties has been made by other researchers. For example, Jakubowicz (2000) argues for the relevance of the notion of *syntactic complexity*, namely that (a) constructions requiring the integration of syntactic knowledge and knowledge from other domains are more complex than constructions requiring syntactic knowledge only, and (b) a syntactic operation is less complex if it is obligatorily required in every sentence; it is more complex if it is present only in some sentences because of semantic or pragmatic choices. Investigating the interface between syntax and discourse necessarily requires going beyond "narrow syntax". Avrutin (1999, 2002) goes a step further and regards "discourse" as "a *computational system* (my emphasis) that operates on non-syntactic symbols and is responsible for establishing referential dependencies, encoding concepts such as 'old' and 'new' information, determining topics, introducing discourse presuppositions, etc...." (Avrutin 2002:1).

The situation of the English speakers of L2 Italian is clear: referential pronouns in Italian qualify as complex, since they demand the simultaneous mastery of both morphosyntactic properties and discourse conditions; in contrast, referential subject pronouns in English are less complex because they are not conditioned by discourse factors. It follows that residual L1 influence leading to optionality in L2 grammars is NOT expected to apply in all cases, but only when the L1 instantiates the most 'economical' option. So English affects Italian in this respect, but in the reverse case of Italian near-native speakers of English one would NOT expect L1 Italian to exert residual influence on L2 English. It is therefore more accurate to say that crosslinguistic influence may take place unidirectionally, from less complex to more complex grammars, whenever two coexisting grammars are in conflict with respect to syntactic complexity. The results from the study on L1 syntactic attrition indeed show that it can be the L2 to affect the L1, if the L2 instantiates the less complex option.

6.   Interpreting optionality: Representational vs. processing accounts

The argument so far has been that residual optionality in the production and comprehension of L2 near-native speakers and L1 speakers under attrition from a second language may be caused by one of the bilingual speaker's syntactic systems – the most economical – affecting the knowledge representations in the other system. In L2 acquisition, this residual influence prevents the complete acquisition of constraints at the syntax-discourse interface. In L1 attrition, this influence causes a change in these constraints. In neither case is 'narrow syntax' directly affected. This account may be termed 'representational' because it assumes effects internal to the speaker's grammatical competence.

It is possible, however, to interpret these patterns in a different way. The question to be addressed is this: are interface problems internal to the learner's representation of syntactic knowledge, or are they external to these representations and created by computational difficulties in integrating knowledge from different domains?

Notice that the notion of 'interface' is ambiguous: regardless of the theoretical standpoint one takes, interface conditions seem to involve extra-syntactic factors and the speaker's ability to coordinate different types of knowledge. Thus, the L2 data illustrated above are compatible with the representational account, but are also compatible with a different assumption: namely, that interface problems may not be due to the persisting effects of the L1 on knowl-

edge representations, but rather to some specific difficulty posed by interfaces. Such difficulty may be due to inadequate processing abilities in coordinating and integrating different types of knowledge. Thus the problems with overt subject pronouns described above may stem from the L2 speakers' not consistently having the computational resources necessary to coordinate the use of an overt pronoun with the introduction of a new or contrastive topic. This alternative, which may be termed the 'processing account', is thus independent of the issue of crosslinguistic influence.

The idea that at least some phenomena in L2 end-states may be attributable to inadequate coordination of different types of knowledge has been gaining ground in L2 research. L2 studies on other potentially problematic interfaces (e.g. the syntax-morphology interface Lardiere 1998; Prévost & White 2000) suggest that persistent (or even potentially permanent) morphological problems in the endstate grammar may be 'surface' problems related to the mapping of abstract syntactic knowledge onto the correct morphological exponents. The fact that learners' problems tend to be with missing inflection, as opposed to wrong inflection, suggests the existence of computation problems with the integration between syntactic and morphological knowledge, leading to the optional use of 'default' underspecified forms.

Furthermore, there is independent evidence from psycholinguistic and neurolinguistic research that processing abilities in L2 speakers are different from those of monolingual speakers. An example is provided by Kilborn (1992), who investigated German near-native speakers of English. His experiments focused on the on-line integration of different types of grammatical information in comprehension. Word-monitoring tasks were used in normal and noise conditions.[14] Kilborn's results show that the performance of near-natives in normal listening conditions is similar to that of natives in noise conditions, indicating a failure to integrate syntactic and semantic information as rapidly as monolingual L1 speakers. Once again, this study points to an interface domain as a prime locus of differences between native and near-native knowledge.

More recently, Felser et al. (2003) suggested that adult L2 learners employ qualitatively different parsing strategies from native speakers (both young and mature). In particular, they tend not to use universal, least-effort strategies based on phrase structure; instead, they attempt a more direct mapping of surface form to interpretation. Faced with ambiguous relative clauses with complex antecedents, such as (25a, b), advanced L2 speakers use different disambiguation strategies from native speakers.

(27) a. Someone shot the servant of the actress who was on the balcony
      b. Everyone liked the actress with the servant who was always smiling

Felser et al. argue that L2 learners do not rely on phrase-structure information to the same extent as both young and mature native speakers. Instead, they attempt more direct form-function mappings, indicating a failure to integrate phrase-structure and lexical-semantic information and, crucially, a lack of automaticity of syntactic processing.

Another window into L2 speakers' processing abilities has been opened by recent cognitive neuroscience research making use of brain-imaging techniques, such as functional magnetic resonance imaging (fMRI) and electrophysiological measurements, such as event-related potentials (ERPs). While this type of research is still very new and should be interpreted cautiously, it has nevetherless revealed potentially important new lines of research. The evidence provided by these studies is completely independent of speaker's voluntary performance, and can be therefore regarded as complementary to behavioural evidence resulting from linguistic and psycholinguistic research.[15]

Summarizing a complex picture (see Hahne & Friederici 2001 for an overview), ERP studies point to:

a. A reduced automaticity of the phrase structure component: this is signaled by the absence of early left anterior negativity (LAN) in non-native speakers, compared to native speakers.[16]
b. Quantitative differences between native and non-native speakers with respect to semantic processing: the telling piece of evidence in this case is a delayed or more pronounced N400 component.
c. Qualitative differences between native and non-native speakers with respect to syntactic integration, evidenced by the absence of a P600 component in low-proficiency learners.

Studies employing FMRI suggest that, overall, different brain regions subserve language processing in L1 and L2.[17] While this by itself is not necessarily an indication of different processing strategies, some studies also indicate that proficiency level and age of first exposure significantly affect activation patterns only for *grammatical* (i.e. syntactic) processing; the effects of these variables are reduced or absent for lexical or semantic processing. L2 processing elicits more inter-subject variability, as indicated by more diffuse activation patterns in L2 speakers. Finally, L2 grammatical (but not semantic) processing involves stronger and more extensive brain activation compared to L1 processing, even when there are no behavioral differences (Wartenburger et al. 2003; Meyer et

al. 2003). This pattern may be taken as an indication of 'greater effort' involved in L2 processing, even in highly proficient L2 speakers.

Taken together, these results are consistent with the behavioural studies which point to a reduced automaticity of grammatical, as opposed to lexical or semantic, processing. Future neuroimaging studies will need to analyse the syntactic component further, separating purely computational aspects from interface aspects. Further research will also need to look at different kinds of integration processes, separating lexical from discourse/pragmatic information. However, it is already possible to extract a message from these studies: there are differences between native and non-native speakers in terms of processing abilities, and these differences selectively concern the access to and the integration of the syntactic component and the coordination of multiple sources of knowledge.

To return to the question of whether optionality in L2 near-native grammars is due to representational or processing problems, one can formulate the following working hypothesis. If the efficiency of L2 syntactic processing is suboptimal, L2 speakers' ability to integrate syntactic knowledge with information from different domains is likely to be sub-optimal too and may fail with significantly more frequency than in L1 speakers. When integration fails, speakers may resort to default strategies, such as the use of overt subject pronouns and the positioning of subjects in preverbal position.[18] Cross-linguistic influence from the L1 may be a reinforcing factor, but it loses its privileged status as the only cause of these phenomena.[19]

7.   Usage and exposure as critical variables

Does the processing account hold for L1 attrition? The assumption in this case would be that syntactic processing in the native language becomes less than optimal as a result of prolonged exposure to a second language. One factor that tends to be systematically underestimated, but needs to be considered in this respect is the role of 'practice', i.e. continuous exposure to input beyond a certain threshold, and continuous active use. What L2 near-native speakers and L1 speakers under attrition have in common is the fact that their total exposure to the language is reduced compared to that of monolingual speakers: in the case of L2 speakers, because they started the process of L2 acquisition in adulthood; in the case of L1 speakers under attrition, because they stopped being exposed to the L1 continuously. The same argument could apply to L1 bilingual acquirers, because they are exposed to two languages simultaneously

and therefore the quantity of input received in each language is – even in the ideal case of perfectly balanced input – half the input to which monolingual children are exposed.

It is possible that quantitatively reduced input may determine a drastic decrease in the number of opportunities for coordinating different information types in communication, and may therefore result in an efficiency loss for these processing abilities. Besides quantitative differences, the input these bilinguals receive may also be qualitatively different from the input in a typical monolingual environment. L2 speakers, especially if they live in country where their own language is spoken, may use the L2 in interactions with other L2 speakers (colleagues, spouses, children) or with L1 speakers under attrition who produce the same non-native forms. Similarly, L1 speakers may hear their native language spoken by other L1 speakers under attrition and by L2 speakers. These speakers' optionality is therefore reinforced by optionality in the input.[20]

To recapitulate, lack of automaticity in syntactic processing may, at least partly, be due to insufficient practice in, and exposure to, a language. If the syntactic component is less automatic in bilingual processing, the integration of different types of knowledge may be computationally more costly. Thus, the parallels among these groups of bilingual speakers suggest that sustained exposure to input may be necessary both for *acquiring and maintaining* an efficient syntactic system.

8. Conclusions

This paper has presented evidence of residual optionality and persistent L1 influence in near-native L2 grammars whose locus seems to be the interface between syntactic and discourse/pragmatic knowledge. Similar patterns of optionality and (asymmetric) crosslinguistic influence are also found in the domain of non-pathological individual L1 attrition. There are two possible explanations for these patterns: one involves underspecification at the level of knowledge representations, with 'soft' interface constraints being the target of indeterminacy; the other involves processing difficulties related to the integration of different types of knowledge. Behavioural and neuropsychological evidence suggest that syntactic processes are less automatic in L2 speakers than in L1 speakers, which in turn may increase integration difficulties. Future research will tell whether only one of these accounts is the correct one, or whether perhaps both are necessary to explain these complex patterns of linguistic behaviour in bilingual speakers.

Notes

1. Here and throughout the paper the focus is on adult L2 acquisition. Child L2 acquisition (i.e. the acquisition of a second language between the ages of 2 and, roughly, puberty) is not considered, although there is evidence that grammatical attainment at this age is generally higher than in adult learners (see Lakshmanan 1995).

2. There is still considerable controversy around the question of whether a language-specific cognitive faculty ("Universal Grammar") constrains adult second language acquisition. On the one hand, research has shown that adult learners can attain complex knowledge of the L2 that goes beyond the input received (White 2003): in this respect, there is a 'poverty of the stimulus' argument for both first language acquisition (see Pinker 1989, among others) and second language acquisition. On the other hand, there are undeniable differences in the outcome of first and second language acquisition, which would seem hard to reconcile with the idea that the same constraints are at work in both cases. It is important to consider the possibility that these differences may be due to other (perhaps domain-general) cognitive faculties and not to Universal Grammar *per se*.

3. Pidgins and creoles also exemplify how natural language grammars are governed by the same constraints that shape language development in other domains (see DeGraff 1999 and the articles therein).

4. I leave aside here the difference between near-native grammars and the vastly more common *non*-native 'fossilized' endstates, which is important (Long 2003; White 2003) but not relevant to the discussion in this paper.

5. Experimental data in Italian were collected by Filiaci (2003), Tsimpli et al. (2004), and Belletti, Bennati and Sorace (2005). The reader is referred to these works for detailed statistical results. Different types of controlled production and comprehension tasks were used to elicit subject pronouns in a range of contexts. Acceptability judgment tasks are not suitable for testing these constructions, since sentences would have to be presented in context. Participants were screened for near-nativeness using a modified version of the test developed by White and Genesee (1996).

6. The fact that both options – null and overt pronouns – are possible in different contexts means that errors in this domain do not give rise to strong ungrammaticality, and native speakers' intuitions are best treated as preferences, rather than categorical behaviors. In fact, the lack of categoricalness in this domain is predicted by the hypothesis pursued in this paper, namely that pronominal choice in Italian is at the interface between syntax and discourse.

7. It is often assumed that unaccusative verbs freely allow postverbal subjects in an 'unmarked' (i.e. unfocused) context as a result of their single argument being generated in the direct object position (see Belletti 1988, 2002). However, Pinto (1997) argues that what determines the possibility of subject-verb 'inversion' in Italian is the presence of a [+Locative] feature either in the argument structure of the verb, or in the predicate in which the verb appears. For this reason, some unergative verbs allow postverbal subjects with a locative expression in an all-focus context (*In questa stanza ha dormito il re di Spagna* 'In this room slept the king of Spain'); on the other hand, some unaccusative verbs that do not denote a telic change

are unnatural with a postverbal subject (?*_Dopo la notizia è impallidita mia madre_ 'After the news got pale my mother).

8. This is also evidenced from the fact that near-native speakers of Italian have knowledge of some of the syntactic correlates of the [+Null Subject] setting, such as the lack of subject-object asymmetries in complex extractions from Wh-islands (see Filiaci 2003; Tsimpli et al. 2004 for details).

9. It remains to be determined whether different types of interface (i.e. the syntax-lexicon interface, the syntax-semantics interface, the syntax-phonology interface) pose different challenges to learners from those of the syntax-discourse interface.

10. It is interesting that sometimes group results may obscure patterns of divergence in individual subjects. Birdsong's replication of Coppieters' study is a case in point: there was great variation in terms of individual performance, with some of the subjects performing in a convergent way and others (most, in fact) performing in a divergent way compared to native speakers. These findings are not uncommon in sociolinguistics either. A classic discussion of individual variability can be found in Macaulay (1978) and Dorian (1994).

11. It is worth noting that variation with respect to residual V2 is found among less educated native English speakers.

12. Clitic climbing is subject to dialectal variation: in general, it is not common in northern Italian varieties and is more widespread in southern varieties, although one finds considerable idiolectal variation within these two main areas.

13. The syntax-discourse interface has also been identified as a domain particularly vulnerable to instability and crosslinguistic influence in bilingual L1 acquisition (i.e. the simultaneous acquisition of two languages from birth). Müller and Hulk (2001) maintain that there are two constraints on crosslinguistic influence in bilingual language development: first, there must be structural correspondence between the bilingual child's two languages with respect to a particular area of grammar; second, the area(s) in question must require the interfacing between syntax and pragmatics. For further theoretical and empirical refinements of this hypothesis, see Serratrice, Sorace and Paoli (2004).

14. A word monitoring task requires participants to press a button whenever they hear a particular target word in a recorded text. Reaction times are measured, enabling the researcher to assess, for example, whether word recognition is affected by the semantic or syntactic ambiguity of the text.

15. It should be noted, however, that to date there are no studies specifically focused on near-native speakers (with the exception of Meyer et al. 2003), and that some studies do not control adequately for L2 proficiency.

16. Event-related potentials are small voltage changes in the electroencephalogram, which reflect the activity of a large number of neurons in response to given stimuli. An ERP consists of positive and negative voltage peaks, called "components", which vary in polarity, latency and distribution on the scalp. ERP effects refer to changes in the amplitude and latency of components as a function of the factors manipulated in the experiment. The main advantage of this method is its excellent temporal resolution, which allows the study of language processing on the temporal scale on which it takes place.

17. FMRI is a hemodynamic method that detects increases in the oxygenated blood flow to particular regions of the brain when they become activated in response to given stimuli. It has very good spatial resolution but is not best suited to the study of real-time processing, since the metabolic responses it measures occur at some temporal distance from the presentation of the relevant stimuli.

18. Indeed, it appears that these difficulties are resolved in ways that betray the influence of universal factors. In the cases presented here, optionality has asymmetric effects and favors the retention and occasional surfacing of unmarked options, which are subject to fewer constraints; this is consistent with typological trends (see e.g. Bresnan 2000 on pidgin formation).

19. There is in fact suggestive evidence from studies of pronominal usage in speakers who do not know English. Serratrice (2004) presents data from older monolingual Italian children who substitute overt subjects for null subjects significantly more often than adult native speakers. Bini (1993) shows that the same phenomenon is found in the spontaneous production of Spanish learners of Italian. In both these cases, crosslinguistic influence from a more economical grammar cannot be an explanation. Rather, the relevant factor seems to be a specific difficulty with the syntax-pragmatics conditions on the distribution of subject pronouns, which causes developmental delays and triggers the use of overt subjects as a default form.

20. The effects of optionality in the input may be compounded for speakers exposed to both standard and non-standard dialects simultaneously.

References

Altenberg, Evelyn P. (1991). "Assessing First Language Vulnerability to Attrition". In Herbert W. Seliger & Robert M. Vago (Eds.), *First Language Attrition* (pp. 189–206).

Avrutin, Sergey (1999). *Development of the Syntax-Discourse Interface*. Dordrecht: Kluwer.

Avrutin, Sergey (2002). "Beyond Narrow Syntax". Ms., University of Utrecht.

Belletti, Adriana (1988). "The Case of Unaccusatives". *Linguistic Inquiry, 19*, 1–34.

Belletti, Adriana (2002). "Aspects of the Low IP Area". In Luigi Rizzi (Ed.), *The Structure of IP and CP. The Cartography of Syntactic Structures*, Vol. 2. Oxford: Oxford University Press.

Belletti, Adriana (2003). "Extended Doubling and the VP Periphery". Ms., University of Siena.

Belletti, Adriana, Elisa Bennati, & Antonella Sorace (2005). "Revisiting the Null Subject Parameter from an L2 developmental perspective". Paper presented at the *XXXI Incontro di Grammatica Generativa*, Rome, February 2005.

Bini, Milena (1993). "La Adquisicíon del Italiano: Mas Allá de las Propiedades Sintácticas del Parámetro *Pro-Drop*". In Juana Liceras (Ed.), *La Linguistica y el Analisis de los Sistemas no Nativos* (pp. 126–139). Ottawa: Doverhouse.

Birdsong, David (1992). "Ultimate Attainment in Second Language Acquisition". *Language, 68*, 706–755.

Borer, Hagit (1996). "Access to Universal Grammar: The Real Issues". *Brain and Behavioral Sciences, 19*, 718–720.

Bresnan, Joan (2000). "The Emergence of the Unmarked Pronoun". In Géraldine Legendre, Jane Grimshaw, & Sten Vikner (Eds.), *Optimality-Theoretic Syntax* (pp. 113–142). Cambridge, MA: MIT Press.

Burzio, Luigi (1986). *Italian Syntax: A Government-Binding Approach.* Dordrecht: Foris.

Coppieters, Réné (1986). "Competence Differences Between Native and Near-Native Speakers". *Language, 63*, 544–573.

DeGraff, Michel (1999). *Language Creation and Language Change. Creolization, Diachrony, and Development.* Cambridge, MA: MIT Press.

Dorian, N. (1994). "Varieties of Variation in a Very Small Place: Social Homogeneity, Prestige Norms, and Linguistic Variation". *Language, 70*, 631–696.

Felser, Claudia, Leah Roberts, Theodore Marinis, & Rebecca Gross (2003). "The Processing of Ambiguous Sentences by First and Second Language Learners of English". *Applied Psycholinguistics, 24*, 453–489.

Filiaci, Francesca (2003). The Acquisition of Null and Overt Subjects by English- Near-Native Speakers of Italian. MSc dissertation, University of Edinburgh.

Giorgi, Alessandra & Fabio Pianesi (1998). *Tense and Aspect: from Semantics to Morphosyntax.* New York: Oxford University Press.

Grimshaw, Jane & Vieri Samek-Lodovici (1998). "Optimal Subjects and Subject Universals". In Pilar Barbosa, D. Fox, Paul Hangstrom, Martha McGinnis, & David Pesetsky (Eds.), *Is the Best Good Enough? Optimality and Competition in Syntax* (pp. 193–219). Cambridge, MA: MIT Press.

Gurel, Ayse (2002). Linguistic Characteristics of Second Language Acquisition and First Language Attrition: Turkish Overt vs. Null Pronouns. Ph.D. dissertation, McGill University.

Hahne, Anja & Angela Friederici (2001). "Processing a Second Language: Late Learners' Comprehension Mechanisms as Revealed by Event-Related Brain Potentials". *Bilingualism: Language and Cognition, 4*, 123–141.

Hawkins, Roger (2001). "The Theoretical Significance of Universal Grammar in Second Language Acquisition". *Second Language Research, 17*, 345–367.

Hopp, Holger (2004). "Syntactic and Interface Knowledge in Advanced and Near-Native Interlanguage Grammars". In Susan H. Foster-Cohen, Michael Sharwood Smith, Antonella Sorace, & Mitsuhiko Ota (Eds.), *Eurosla Yearbook* (pp. 4.67–94). Amsterdam: John Benjamins.

Jakubowicz, Celia (2000). "Functional Categories in (Ab)Normal Language Acquisition". Ms., Université Paris 5.

Kellerman, Eric (1987). Aspects of Transferability in Second Language Acquisition. Ph.D. dissertation, University of Nijmegen.

Kilborn, Kerry (1992). "On-Line Integration of Grammatical Information in a Second Language". In Roy. J. Harris (Ed.), *Cognitive Processing in Bilinguals* (pp. 337–350). Amsterdam: Elsevier Science.

Klein, Wolfgang & Clive Perdue (1997). "The Basic Variety (Or: Couldn't Natural Languages Be Much Simpler?)". *Second Language Research, 13*, 301–347.

Lakshmanan, Usha (1995). "Child Second Language Acquisition of Syntax". *Studies in Second Language Acquisition, 17*, 301–329.

Lardiere, Donna (1998). "Case and Tense in the 'Fossilized' Steady State". *Second Language Research, 14*, 1–26.

Leonini, Chiara & Adriana Belletti (2004). "Subject Inversion in L2 Italian". In Susan H. Foster-Cohen, Michael Sharwood Smith, Antonella Sorace, & Mitsuhiko Ota (Eds.), *Eurosla Yearbook* (pp. 4.95–118). Amsterdam: John Benjamins.

Lightfoot, David (1999). *The Development of Language*. Oxford: Blackwell.

Long, Michael (2003). "Stabilization and Fossilization in Interlanguage Development". In Catherine Doughty & Michael Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 487–536). Oxford: Blackwell.

Macaulay, R. K. S. (1978). "Variation and Consistency in Glaswegian English". In Peter Trudgill (Ed.), *Sociolinguistic Patterns in British English* (pp. 132–143). London: Arnold.

Meisel, Juergen (2001). "From Bilingual Language Acquisition to Theories of Diachronic Change". Paper presented at the *3rd International Symposium on Bilingualism*, Bristol, April 2001.

Meyer, Martin, Nigel Goddard, Enrico Simonotto, Adam McNamara, A. Rayna Azuma, Susanna Flett, & Antonella Sorace (2003). "Differential Brain Responses to L1 and L2 in Near-Native L2 Speakers". Poster presented at *Human Brain Mapping* 2003, New York.

Montrul, Silvina (2002). "Incomplete Acquisition and Attrition of Spanish Tense/Aspect Distinctions in Adult Bilinguals." *Bilingualism: Language and Cognition, 5*, 39–68.

Montrul, Silvina (2004). "Subject and Object Expression in Spanish Heritage Speakers: A Case of Morphosyntactic Convergence". *Bilingualism: Language and Cognition, 7*, 125–142.

Montrul, Silvina & Roumyana Slabakova (2003). "Competence Similarities Between Native and Near-Native Speakers: An Investigation of the Preterite/Imperfect Contrast in Spanish". *Studies in Second Language Acquisition, 25*, 351–398.

Müller, Natascha & Aafke Hulk (2001). "Cross-linguistic Influence in Bilingual Language Acquisition: Italian and French as Recipient Languages". *Bilingualism: Language and Cognition, 4*, 1–22.

Oka, T. (1996). "Scrambling in Japanese and English". *MIT Working Papers in Linguistics, 29*, 361–388. Cambridge, MA: MIT Press.

Pinker, Steven (1989). *Learnability and Cognition*. Cambridge, MA: MIT Press.

Pinto, Manuela (1997). Licensing and Interpretation of Inverted Subjects in Italian. Ph.D. dissertation, University of Utrecht.

Polinsky, Maria (1995). "American Russian: Language Loss Meets Language Acquisition". In W. Browne, E. Dornish, N. Kondrashova, & D. Zec (Eds.), *Annual Workshop on Formal Approaches to Slavic Linguistics* (pp. 371–406). Ann Arbor: Michigan Slavic Publications.

Polinsky, Maria (in press). "Incomplete Acquisition: American Russian". To appear in *Journal of Slavic Linguistics*.

Prévost, Philippe & Lydia White (2000). "Missing Surface Inflection or Impairment in Second Language Acquisition? Evidence from Tense and Agreement". *Second Language Research, 16*, 103–134.

Rizzi, Luigi (1982). *Italian Syntax*. Dordrecht: Foris.

Robertson, Daniel & Antonella Sorace (1999). "Losing the V2 Constraint". In Elaine Klein & Gita Martohardjono (Eds.), *The Development of Second Language Grammars. A Generative Approach* (pp. 317–361). Amsterdam: John Benjamins.

Seliger, Herbert (1991). "Language Attrition, Reduced Redundancy, and Creativity". In Herbert W. Seliger & Robert Vago (Eds.), *First Language Attrition* (pp. 227–240). Cambridge: Cambridge University Press.

Serratrice, Ludovica (2004). "Anaphoric Interpretation of Null and Overt Pronominal Subjects in Bilingual and Monolingual Italian Acquisition". Ms., University of Manchester.

Serratrice, Ludovica & Antonella Sorace (2003). "Overt and Null Subjects in Monolingual and Bilingual Italian Acquisition". *Proceedings of BUCLD, 27*, 739–750.

Serratrice, Ludovica, Antonella Sorace, & Sandra Paoli (2004). "Transfer at the Syntax-Pragmatics Interface: Subjects and Objects in Italian-English Bilingual and Monolingual Acquisition". *Bilingualism: Language and Cognition, 7*, 183–206.

Sorace, Antonella (1993). "Incomplete and Divergent Representations of Unaccusativity in Non-Native Grammars of Italian". *Second Language Research, 9*, 22–48.

Sorace, Antonella (1999). "Initial States, End-States, and Residual Optionality in L2 Acquisition". *Proceedings of BUCLD, 23*, 666–674.

Sorace, Antonella (2000a). "Differential Effects of Attrition in the L1 Syntax of Near Native L2 Speakers". *Proceedings of BUCLD, 24*, 719–725.

Sorace, Antonella (2000b). "Syntactic Optionality in L2 Acquisition". *Second Language Research, 6*, 93–102.

Sorace, Antonella (2003). "Near-Nativeness". In Catherine Doughty & Micheal Long (Eds.), *Handbook of Second Language Acquisition* (pp. 130–151). Oxford: Blackwell.

Sorace, Antonella (to appear). "Gradedness and Optionality. Experimental Investigations on Stable and Developing Grammars". To appear in Gisbert Fanselow & Caroline Féry (Eds.), *Gradedness in Grammars*. Oxford: Oxford University Press.

Sorace, Antonella & Frank Keller (2005). "Gradience in Linguistic data". *Lingua*.

Tsimpli, Ianthi, Antonella Sorace, Caroline Heycock, & Francesca Filiaci (2004). "First Language Attrition and Syntactic Subjects: A Study of Greek and Italian Near-Native Speakers of English". *International Journal of Bilingualism*.

Tsimpli, Ianthi, Antonella Sorace, Caroline Heycock, Francesca Filiaci, & Maria Bouba (2003). "Subjects in L1 Attrition: Evidence from Greek and Italian Near-Native Speakers of English". *Proceedings of BUCLD, 27*, 787–797.

Wartenburger, I., H. R. Heekeren, J. Abutalebi, S. F. Cappa, A. Villringer, & D. Perani (2003). "Early Setting of Grammatical Processing in the Bilingual Brain". *Neuron, 37*, 1–20.

Weltens, Bert, Kees de Bot, & Theo van Els (Eds.). (1986). *Language Attrition in Progress*. Dordrecht: Foris.

White, Lydia & Fred Genesee (1996). "How Native is Near-Native? The Issue of Ultimate Attainment in Adult Second Language Acquisition". *Second Language Research, 12*, 233–265.

White, Lydia (2003). *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

Syntactic variation and spoken language*

Jenny Cheshire
Queen Mary, University of London

1. Introduction

Linguists who analyse spoken language have often commented on the challenge that it poses for conventional analyses of syntactic structure. An early comment of this kind was from Crystal (1976:166), whose analysis of clause structure in spontaneous conversational English led him to claim that the linguistic organization of this variety of English had been "fundamentally misconceived". Crystal attributed our lack of understanding partly to the absence of data and partly to the uncritical application of traditional paradigms of enquiry. We can no longer claim an absence of data, now that electronic corpora of transcribed speech are readily available; but it is still not always recognised that we cannot easily understand the nature of spontaneous spoken language in terms of the standard linguistic descriptors (Milroy 2001:270). In this paper I discuss two fundamental characteristics of spoken language that do not fit well with traditional paradigms and that in my opinion are relevant to a potential alliance between the fields of variationist linguistics and generative linguistics.

The first characteristic is the high proportion of prefabricated expressions that occur in spontaneous unplanned speech. This is relevant to generative approaches to syntactic variation because what may appear to be syntactic structures to be explained within the framework of the internal grammar may be better seen as prefabricated expressions with little flexibility. They would be better accounted for, therefore, within another component of the grammar, or perhaps at an interface level. Their relevance for variationist linguistics is that what we may wish to analyse as a syntactic variant may be more appropriately seen as a lexical item. I will discuss some examples in Section 4.

The second characteristic of spoken language to be considered here is the pervasiveness of affective meanings. These can be difficult to accommodate within conventional generative and variationist frameworks alike, but they need to be taken into account since the expression of affective meanings may influence the extent to which variation arises in syntax. I will give some examples in Section 5. The examples in Sections 4 and 5 will also illustrate some of the difficulties that arise if different syntactic structures are treated in variationist research as surface realisations of one and the same variable.

I will argue that a discourse-oriented analysis is a necessary complement to the analysis of syntactic variation, firstly because it can help to identify structures that are prefabricated and secondly because it can show how speakers use syntactic structures in spoken interaction. This in turn may give some insights into when, and why, variation occurs. I begin however by noting the lack of attention to the structure of spoken language within both the generative and the variationist traditions.

2.   The generativist approach and spoken language

Generative theory aims to characterise our innate knowledge of language structure. It is assumed that this knowledge is identical for the individual speakers of a language (indeed, at a more general level, for speakers of all languages) and that appropriate data can come from a single speaker's intuitive judgements about the grammaticality of constructions. The theory has not been concerned with performed, externalised language, so it has not been necessary to consider either the syntactic organisation of spoken language or the potential differences between spoken and written varieties of language. However, now that generativists have begun to work with data arising from variationist research it is impossible not to consider these questions, for the data used in the study of variation and change come from what in the generative tradition is considered externalised language. Variationists prioritise spoken rather than written language (Schneider 2002:67) and there has been a tradition of analysing wherever possible the most spontaneous unplanned 'vernacular' speech, where speakers pay the minimum attention to its production and the Observer's Paradox is least in evidence (Labov 1970). Thus for 'internal' and 'external' approaches to be mutually supportive rather than conflicting, as Chomsky has recently argued that they should be (Chomsky 1999:34), it is necessary to consider the extent to which the generative model of syntax is appropriate for the analysis

of spoken language as well as for the analysis of data obtained from elicited intuitions.

There are arguments in favour of seeing the generative model as appropriate for externalised language as well as for data obtained from intuitions. For example, the intuitions of speakers of languages with a strong written tradition, such as English, are likely to be heavily influenced by the written language (Brown & Yule 1983). Furthermore, some corpus linguists have argued that there is a shared common core between the syntax of spoken and written language (for example, Leech 2000). This would suggest that data obtained from intuitions may fit well with the structure of externalised language, both written and spoken. It can also be claimed that generative theory has developed in part from the intuitions of speakers of languages that do not have a written form. If we can accept that speakers have an intuitive knowledge of the structures they produce when speaking, this is a further reason for claiming that the model is applicable to spoken language.

However, several researchers who have analysed corpora of spoken language claim that the structures of spoken language differ both from data obtained from intuitions and from the syntax of planned written language. For example, spoken English – but not planned written English – contains utterances such as (1) and (2) below, uttered within a single intonation contour (see Cheshire 1999). Here, two clauses appear to be fused: in (1) these are *that's really what Professor Galbraith was talking about* and *what Professor Galbraith was talking about is that there's a huge knock-on effect:*

(1)   That's really what Professor Galbraith was talking about is that there's a huge knock-on effect

(2)   It's just a bunch of baloney is what it is

Crystal (1976) gives more complex examples, including some involving intercalation, as in (3). The clauses are set out on separate lines as in Crystal (1976: 158) for ease of description.

(3)   a.   I'm very suspicious of the press generally
      b.   and I can tell you
      c.   because not only I mean that's one case
      d.   that you've given
      e.   but also on their reporting of erm affairs foreign affairs
      f.   because living in Cyprus
      g.   I've seen quite a number of historical events you know

Here the clauses in (3c–e) are subordinate to the clause in (3a) and those in (3f–g) are subordinate to the clause in (3b). Crystal points out that from a semantic viewpoint (3e) relates to both (3a) and (3f): it is half of the reason for the proposition expressed in (3a), along with (3c)and (3d), but it also constitutes a new theme linking (3c) with (3f) and (3g). The syntactic status of (3e) is unclear: there is no main verb and it could, Crystal argues (op. cit.:158), be analysed as a complex adverbial linked via the *because* of (3f) to (3g).

There are several accounts now of syntactic forms that are specific to spoken language. These include, for English, Biber et al. (1994), Miller (in press); for English and Russian, Miller and Weinert (1998) and for French, Blanche-Benveniste (1997). It is not clear how structures such as these could be considered as generated by the grammar, yet many researchers working outside the generative paradigm have argued that they should be included within a model of grammar. Schegloff (1989:143), for example, argues that a biological perspective on language should examine language in the natural environment in which it occurs (in other words, in unplanned conversation) and should entertain the idea that the syntactic structures of spoken language are adaptations to this environment. A growing body of work shows that spoken language has its form partly as a result of interactional factors, such as the turn-taking mechanism (for example, Ochs et al. 1996; Couper-Kuhlen & Selting 2001; and Ford 1993), and that these factors help determine variation. For example, Ford (op. cit.:147–148) finds that variation in the clause position of conditional and temporal adverbial clauses in American English is related to their discourse functions and the conversational context. They are more frequent in clause-initial position in extended spans of talk, where speakers have negotiated a special right to the floor and where there are more opportunities for the discourse-structuring functions of the clauses to be realised. Mondorf (2000) adds a sociolinguistic perspective: in the London-Lund corpus of spoken British English male speakers tend to position adverbial clauses initially, whereas female speakers place them more frequently in clause-final position. Mondorf explains these preferences in terms of the different orientations of male and female speakers to information management and epistemic grounding. Levinson (1983:97) has argued more generally that social principles for co-operative interaction, such as those shown in politeness strategies, have a pervasive effect on language structure.

It has been claimed, then, from several different quarters, that interactional and social factors can constrain both the form of spoken syntax and aspects of syntactic variation. It is not yet clear to what extent the structure of spoken syntax can be explained as the result of performance mechanisms that do not

need to be accounted for within the internal grammar (as it is conceived by generativists), but it becomes difficult to avoid the question when data from spoken language are used to develop generative theory.

3.   The variationist approach and spoken language

Since researchers working in the variationist tradition have always worked with the data of 'externalised' language, they might be expected to have paid attention to the characteristics of spoken syntax. However here too the nature of spoken language has been largely neglected, albeit for different reasons.

One reason is that the linguistic variable was originally conceived for the analysis of phonological variation. Variants should be semantically equivalent: in other words, they should be alternative ways of 'saying the same thing' (Chambers & Trudgill 1998: 50). Semantic equivalence can be easily established for phonological variables, where the form-meaning relationship is at its most arbitrary, but there has been much controversy about whether it can also be established for syntactic variation. The issues were much discussed during the 1970s and 1980s (for example, Lavandera 1978; Cheshire 1987; Levinson 1988; Romaine 1980; Weiner & Labov 1983) and debate has continued since then (see, with reference to French, Blanche-Benveniste 1997; Coveney 1997; Gadet 1997 and for general discussions Cheshire, Kerswill, & Williams 2005; Cornips & Corrigan 2005; Coveney 2002; Milroy & Gordon 2003). A tacit consensus seems to be that the condition of strict semantic equivalence can be relaxed for syntactic variables, so that a variable can be set up on the basis of an equivalence in discourse function (Dines 1980; Coupland 1983). For example, the five forms in (4) to (8) below (from Romaine 1984: 426) can all be considered ways of 'saying the same thing' in that they all have the same communicative intent or discourse function (the speaker would like the addressee to close the window):

(4)   It's cold in here

(5)   I'm cold

(6)   Are you cold?

(7)   Would you close the window?

(8)   Close the window

The problem however is that we are now dealing with variation that is constrained by pragmatic factors rather than by the grammar. It is not clear to

what extent forms that we might wish to consider as examples of syntactic variation are always motivated by pragmatic factors, and there is no consensus in the field on whether this invalidates their analysis as linguistic variables (see Winford 1996: 188 for further discussion). The issues are no longer much discussed within the variationist literature, but the legacy of the debate accounts in part for the neglect of syntactic variation relative to phonological variation.

A second reason for the neglect of spoken syntax in variationist research is that the methodology does not require a detailed analysis of syntax. Researchers focus on one linguistic variable at a time, extracting tokens from the conversational contexts in which they occur. For example, an analysis of *was/were* variation in an English dialect involves extracting and coding all the tokens of *was* and *were* produced by the speakers participating in the study and then performing a quantitative analysis to determine the linguistic and social constraints on the occurrence of *was* rather than *were* and vice-versa. Once the variable to be analysed has been selected, there is no need to consider the syntax other than to identify which potential internal constraints are to be included in the statistical analysis. Different researchers have chosen for themselves what to include as a potential constraint and decisions are not always based on systematic syntactic grounds (see Henry 2002). Analysts working on *was/were* variation, for example, usually take account of agreement between the subject and the *was/were* form and the polarity of the construction in which the forms occur (in some dialects negation favours *weren't* rather than *wasn't*), but the focus of the analysis remains the *was/were* variable and the procedure does not necessarily require a fuller analysis of the syntactic construction in which *was/were* occurs, nor of the overall discourse structure. Syntactic constructions specific to spoken language, therefore, such as (1) to (3) above, have tended to be overlooked, as have the possible effects of interactional factors.

A related reason for the neglect of spoken syntax in this field is that variationists tend to analyse the same grammatical variables over and over again. This is partly because the favourite variables tend to occur with the high frequencies that are necessary for quantitative analyses. However another, less obvious, reason seems to be that analysts have been influenced by the ideology of the standard. *We was*, for instance, is non-standard relative to standard English *we were* and this has made the form salient not only to prescriptivists and laypeople but also to linguists. In addition, the sociocultural processes involved in standardisation means that the use of non-standard forms correlates with a speaker's social status. Variables of this type are eminently suitable for analysis within the variationist framework since they meet the classic definition of a sociolinguistic variable (a structural unit with two or more variants in-

volved in co-variation with social variables). Variationists have worked almost exclusively on languages that have been heavily standardised, so the potential influence of the standard ideology on the selection of variables for analysis has been high (see Cheshire & Stein 1997 and Milroy 1999 for further discussion). Thus, for English the most frequently analysed variables are morphosyntactic forms where one variant is prescriptively non-standard: they involve subject-verb agreement, negative concord and non-standard negative forms such as *ain't*, as well as various standard and non-standard verb forms. Syntactic variation involving less salient forms have been largely ignored.

Both generativists and variationists, then, have largely neglected the structure of spoken language. Despite the different methodologies that they traditionally adopt – with generativists relying on intuitions and variationists on audio recordings of spoken interaction – each approach has analysed forms abstracted from the interactional context in which they occur. A further similarity is that each approach focuses on abstract linguistic systems, with generativists aiming to characterise the structures of the internal grammar and variationists seeking to understand the properties of variable systems (Milroy & Gordon 2003:8). An essential difference, of course, is that generativists assume that the grammar is identical for all speakers of a language. Their concern, therefore, is to identify which aspects of the observed variation are categorical and how this variation can be accommodated within the grammar. Variationists on the other hand are also concerned with understanding the social embedding of variation, especially insofar as this can explain the processes by which orderly linguistic change occurs. Variation for them is integral to the nature of language and must be incorporated within the model. In principle, therefore, variationists are equally interested in social (external) and internal constraints on variation, though different studies may tend to give more weight to one than the other.[1]

In any event, for both variationists and generativists the analysis of syntactic variation rests on an appropriate identification of the forms that are considered to alternate. I hope to show in what follows that a discourse-oriented approach can help to identify these forms and that this is a necessary complement to generative and variationist analyses.

4.   Prefabricated expressions

4.1   Prefabricated expressions in spoken language

Many researchers working on spoken language have claimed that linguists tend to over-emphasise the creative aspect of language. There is no doubt that we can produce and understand an infinite number of sentences that we have never heard before but, as Bolinger (1975: 297) pointed out, the fact that we *can* do this does not mean that we *do*. It would be counter-productive in spontaneous face-to-face communication to constantly produce brand new sentences and speakers use prefabricated expressions to help them cope with the demands of fast speech production. These expressions include conversational routines with clear social or cultural functions, such as forms conventionally expressing apologies, thanks, compliments or requests (for example the English request formulae *I wonder if I could/ could you possibly/ can I just*), frequent collocations, like *heavy smoker*, *white coffee*, 'construction templates' such as *as far as I* (*can see/know/can make out*), or sentence builders such as *my point is, I'm a great believer in* (see Crystal 1995: 162ff.; Wray 2002).

   Estimates of the proportion of ready-made chunks of unanalysed language in large-scale corpora of spoken language range from 30 per cent (Biber et al. 1999) to 70–90 per cent (see Aijmer 1996: 31). The difference in the estimated proportions reflects the ways in which the chunks are defined. Sometimes researchers rely on subjective identifications of what counts as prefabricated, whilst others give a strict definition on the basis of collocation patterns within a large corpus. For example, Biber et al.'s (1999) analysis of a 40 million word corpus of spoken and written English excludes combinations of less than three words: it therefore excludes recurrent noun and adjective combinations such as *heavy smoker* and recurrent conversational routines like *I'm sorry*. Estimates of the proportion of prefabricated expressions also reflect decisions about how fixed in form an expression must be in order to be considered prefabricated. *How do you do*, for example, is completely frozen and the 'sentence builder' (Crystal 1995: 162) *what I mean is* is capable only of limited alteration (such as *what I really mean is, what I meant to say was*). Some conversational routines have greater flexibility; these include, for example, the compliment formula *I (really) like/love your NP*, where the NP must refer to an item that is culturally approved (Holmes 1995). Aijmer (1996: 217) accounts for the flexibility of certain conversational routines by seeing them as 'mini-grammars' consisting of collocational stems generating a limited set of structures. An example is the expression *to put it another way*: in the London-Lund corpus this could

be described as having a stem generating the related discourse forms *putting it, put it* and *put*, followed by one of four manner adverbials (*this way, like this, another way* and *mildly*). The interrogative *how shall I put it* also occurred in the corpus. Aijmer proposes that conversational routines can be arranged along a continuum from completely fixed forms through semi-fixed forms (e.g. *I'm so/really/very sorry*), frame and slot forms (e.g. *could I have X*) to mini-grammars. It is difficult to accommodate mini-grammars within a formal grammar, however, because their output is so constrained.

Even a strict definition of what constitutes a prefabricated expression gives their proportion within a corpus as 30 per cent: a sufficiently high proportion for their existence to be taken seriously. They raise the question of whether spoken language might be better conceptualised as linear and sequential in structure rather than as hierarchical. The idea is pushed to its limits by Sinclair (1991:68), who predicts that "lexical hordes" will invade the traditional domain of syntax and lead to its eventual demise. Skehan (1998:37) takes a more moderate view, suggesting that the production of speech involves improvising on a clause by clause basis, such that speakers use lexical phrases and lexical sentence stems wherever possible in order to minimize processing demands and only as a last resort generate language that is not part of our memorised lexicon. Even a moderate view, however, suggests that when we are analysing spontaneous spoken language it is important to bear in mind that what may appear to be a syntactic construction may instead be a chunk of ready-made memorised language. I will consider some examples from spoken English in the following section.

4.2 Some prefabricated expressions in spoken English

Independent adverbial phrases in spoken English are a case in point. These appear to be subordinate adverbial clauses in that they are introduced by conjunctions such as *because, when,* or *if*, but there is no main clause. Generative theory does not allow for the possibility of unattached adverbial clauses – understandably, since by definition an adverbial clause is subordinate to a main clause (and they may well be overlooked by researchers, since it is not clear that constructions such as these are accessible to our intuitions). However, both Mondorf (2000) and Ford (1993) noted unattached adverbial clauses in their analyses of adverbial clauses in spoken English. Mondorf reported 6 per cent out of the total number of adverbial clauses (259, out of 4462 clauses) and Ford found 3 per cent out of the total number of temporal adverbial clauses (2 out of 63 temporal clauses). Both authors were able to infer a main clause from the surrounding

linguistic context, but it is not always possible to do so. McCarthy (1998: 79–82) for example notes clauses introduced by *if* and *cos* (a reduced form of *because*) that occur alone and function as main clauses. I found it equally impossible to infer a main clause for some *when* structures that occurred in a corpus of conversations between 12–16-year-old working-class adolescents in Reading, Berkshire (see Cheshire 1982). Unattached phrases introduced by *when* were relatively frequent in my Reading corpus, accounting for 25 per cent (28) of the 105 *when* clauses. I will discuss these phrases in some detail in order to illustrate the problems they can pose for a variationist analysis.

Two examples of the *when* phrases are indicated by the arrows in (9) and (10). They were uttered with level tones on every syllable except the last: this has a falling tone and is slightly drawled. Interestingly, they were used only by the male adolescents.

(9)   (*the boys are talking about one of their teachers, who was married to someone I knew. Jenny* (*me*) *was the fieldworker*)

Nobby:	yeah Miss Threadgold she ain't bad
Rob:	yeah she . she went camping with us
Jenny:	yes he told me she'd been camping
→ Nobby:	when we went camping
Rob:	she's a good laugh
Jenny:	is she?
Nobby:	yeah

(10)   (*the discussion has been about jobs the girls might consider doing when they leave school*)

Jenny:	you have to do horrible jobs if you're a nurse .. all the bed pans
All:	<LAUGHTER>
Jenny:	have you ever been in hospital?
Valerie:	[I have
Christine:	[oh yeah I have
Valerie:	I got run over by a car
Christine:	I fell off a gate backwards <LAUGHS> and I was unconscious
→ Tommy:	oi when I .. when I went in hospital just for a little while …
Valerie:	sshh
Tommy:	cos my sister and my cousin they bent my arm .. they twisted it right round

A variationist analysis of the *when* phrases would seem in principle to be possible, if we assume one variant to be a *when* clause that is clearly subordinate to a main clause (for example, *when we went camping* in *when we went camping we had a great time*) and another variant to be an unattached *when* clause as in (9) or (10).

The first step in a variationist analysis would then be to establish the discourse function of the lone *when* phrases and the conventional, subordinate, *when* clauses, to ensure that they are equivalent in function. One function of conventional initial *when* clauses is explicatory (Ford 1993: 29, 32). Ford found that this was the case when *when* clauses followed a semantically broad term such as *thing* or *then*. In her data the explication occurred within an extended speaker turn. She argued, in fact, that the use of the semantically broad term contributed to the projection of an extended turn. Only four of the lone *when* clauses in the Reading data were explicatory, however. One of these is illustrated in (11): here Rob explains, in answer to a question, how Britt (one of the playground leaders) tries to control her mind. The lone *when* clause does not elaborate a semantically broad term, nor does it project an extended turn, but it does provide a time frame for a specific situation that illustrates Britt's behaviour. In doing so, it clarifies a semantically problematic concept (the idea of controlling your mind) that the emerging discourse has shown to be ambiguous or too vague for present purposes: this was initially unclear to all the participants, as indicated by Rob's *whatever that means* and Nobby's response (*I don't know*) to my question about how this can be done.

(11)   Rob: and Britt she's queer =   = she's trying to learn to control her mind

Nobby: = yeah =

Rob: whatever that means

Jenny: is she?

Rob: [yeah

Nobby: [yeah

Jenny: oh how is she going to what is she doing to con

Nobby: I don't know

→ Rob: when you look at smoke and that you know fire =

Jenny: = yeah

Nobby: she looks at a flame she's . you can look at . she's trying to look at a flame until it burns right out

Jenny: and then w . how does that control your mind?

Rob: I don't know

The four lone *when* phrases with an explicatory function, then, do share at least one of the functions of subordinate *when* clauses.

A further function of subordinate adverbial clauses in initial position is to project an extended turn and present background for material that follows. These characteristics contribute to Ford's view that initial adverbial clauses are pivotal points in the development of talk (op. cit.: 62). The remaining 21 lone *when* phrases in the Reading corpus share these characteristics. In (10), for example, the other speakers interpret Tommy's lone *when* phrase, prefaced by his attention-getting *oi,* as an indication that he intends to take a projected turn; this is shown by Valerie compliantly telling her younger sister to be quiet. Usually, the extended turns are narratives of personal experience; thus, in (10) Tommy went on to tell the story of his stay in hospital. Both explicatory and pivotal lone *when* phrases, then, share some aspects of the interactional function of conventional adverbial *when* clauses. As mentioned above, there is social variation in that the forms without an accompanying main clause are used only by the male adolescents. These forms might seem, then, to be candidates for a variationist analysis, with a sociolinguistic variable consisting of two variants, one a *when* clause with a main clause, the other a lone *when* construction without a co-occurring main clause.

However, this approach would miss an important discourse function of the 21 lone *when* constructions that are pivotal in the development of talk. In every case, the narrative that follows the lone *when* phrase concerns events that are familiar to the other speakers, either because they have heard the story before, or because they participated themselves in the events that are recounted. The narrative is a form of joint reminiscing – a discourse event with an important role in reinforcing group membership (Edwards & Middleton 1986). In the Reading playground conversations these narratives were especially significant in the construction and reinforcement of group friendship patterns amongst the male adolescents. The main function of these lone *when* phrases, in other words, is as a story opener, marking the upcoming story as a shared reminiscence. Female adolescents constructed friendships on a more individual basis, telling stories mainly as monologues. Their different narrative style was reflected in their preferred story opener which, as Table 1 shows, was a temporal subordinate clause, clearly situating the story in the past (for further details see Cheshire 2000).

When the lone *when* phrases are considered in their full interactional context, it becomes clear that they cannot be analysed as variants of conventional initial *when* clauses, since they are not functionally equivalent. They have a specific discourse function as a story opener marking a shared reminiscence. A

Table 1. Story openers in the corpus of narratives

	All-female conversations	All-male conversations	Total
Markers of a shared reminiscence			
Remember when clause	0	1	1
What about that time when	0	2	2
You know when clause	0	1	1
What about X	1	0	1
Pivotal lone *when*-clause	0	5	5
I can't forget that time when	0	1	1
Total markers	**1**	**10**	**11**
Temporal subordinate clauses introduced by			
once	6	0	6
when	5	2	7
the other day	5	0	5
one time	2	0	2
one day	3	0	3
last time	1	0	1
yesterday	1	0	1
Total temp. sub. clauses	**23**	**2**	**25**
Zero	**9**	**34**	**43**
Miscellaneous			
clause *right*	1	7	8
left dislocation	9	3	12
there was X	6	3	9
you know X	2	2	4
you see	3	0	3
see	2	0	2
you should have seen X *mate*	0	1	1
fuck me	0	1	1
he's a bastard mate	0	1	1
oh it's horrible	1	0	1
it wasn't half fun	1	0	1
Total all story openers	**58**	**64**	**122**

variationist analysis could, perhaps, be performed on the range of story openers that are used to introduce sequences of joint reminiscing (those shown as the first group in Table 1), but this of course would hold no interest for the study of syntactic variation.

An analysis that fits better with the data is to see the lone *when* phrases as conversational routines, together with the other story openers marking an upcoming shared reminiscence (such as *what about when, you know when* or

remember when). As we saw earlier, a conversational routine is a sequence of words that appears to have syntactic structure but that is produced and processed as a more or less prefabricated phrase (Aijmer 1996). The *when* of the lone *when* phrases may be a reduced form of the other *when* phrases in this group of story openers marking shared reminiscences.[2]

The lone *when* phrases used as story openers are not, of course, completely fixed in their form: they differ, therefore, from prefabricated phrases such as *how do you do?* and are more productive than the *to put it* expressions mentioned in the previous section. Yet they have more in common with prefabricated lexicalised forms such as these than with completely new clauses that have been generated by the grammar. They consist of a frame (*when* + NP + VP, with the verb in the past tense) with a fixed intonation contour. The past tense form of the verb distinguishes the story openers from the other, less frequent lone *when* phrases with an explicatory function: in (11), for example, the verb *look* is in the present tense. The words that constitute the NP and the VP are repeated from the preceding discourse and this facilitates their function as a way of taking the floor: thus in (9) Nobby's *went camping* echoes the words of the preceding three turns and in (10) Tommy's *in hospital* echoes the question *have you ever been in hospital?*

What might initially appear to be an instance of syntactic variation, then, is more appropriately seen as a conversational routine with an interactional function in turn-taking and a social function in indexing group solidarity (as we have seen, it is used only by the boys, along with other story openers that mark an upcoming shared reminiscence). It is not entirely fixed in form and conforms more to a phrase generated by a 'mini-grammar'.

Other forms used as story openers in the Reading conversations are better analysed as prefabricated expressions than as constructions generated by the grammar. One such form involves verbal *-s*. This of course is usually considered to be an agreement marker in generative analyses of English and in present-day standard English it does indeed appear to have this function, occurring only on present tense verb forms with third singular subjects. In many present-day non-standard varieties however the distribution of verbal *-s* differs. In Norwich, England, for example, it is variably absent with third person subjects (Trudgill 1974); in Reading it is variably present with non-third person subjects and quasi-categorical with third person subjects (Cheshire 1982), as in several other varieties of British and North American English. It is sometimes assumed that speakers have regularised the present tense paradigm in these vernaculars, so that verbal *-s* is an agreement marker in these vernaculars also, but many researchers have identified a wider, diverse range of functions for the

form, perhaps especially in African American English. The functions include marking durative aspect (Pitts 1986; Brewer 1986), habitual aspect (Pitts 1986), variably marking the present tense (Schneider 1983) and marking the historic present (Myhill & Harris 1986). Overviews of research on verbal -*s* in English dialects are given by Clarke (1997) and Godfrey and Tagliamonte (1999). Most of these studies exclude from the envelope of variation story openers or topic introducers such as *you know* in (12) and (13). In the Reading corpus, as elsewhere, *you know* used in this way is invariable, never taking the -*s* suffix, unlike *you know* as a lexical verb, as in (14) and (15):

   (12)   you know that hill down there? I rode down that with no hands on the handlebars

   (13)   you know your mum. . . .you know that bike she had

   (14)   you knows him don't you Nod?

   (15)   he says to me "look here and I see if I knows you"

The story opener then, is a prefabricated expression, like the discourse marker *you know* (which performs a range of conversational functions, including adding liveliness to a conversation and constructing solidarity; see, for example, Holmes 1986; Fox Tree & Schrock 2002:729). It is not certain, however, that all cases of prefabricated expressions have been accounted for in analyses of verbal -*s*. After all, existential constructions and canonical clause constructions are usually analysed side-by-side, despite their different syntactic derivations (see Corrigan 1997 for discussion). Analyses of verbal -*s,* whether generativist or variationist, might benefit from a prior discourse analysis aiming to identify all the prefabricated expressions in which verbal -*s* occurs: this would not only make the analyses more accountable to the data but would also further our understanding of how and why prefabricated expressions develop and their role in grammaticalisation and other kinds of language change.

5.   Affective meanings in spoken language

The prevalence of emotive expressions in conversation is well known (see for example Biber et al. 1999:958), as is the fact that for speakers the communication and construction of affective meaning is as important as the communication of referential meaning. Here I will simply consider the relevance of this fundamental characteristic of language for determining the extent to which syntactic variation may arise in discourse.

As the main example I will consider the pronoun tags that occur in several northern varieties of British English. They are illustrated in (16) and (17) below, indicated by the arrows, with the tag in italics. The extracts are taken from an analysis of variation in the speech of 14–15 year old working-class and middle-class adolescents in Hull, England (Cheshire, Kerswill, & Williams 1999). In our data these pronoun tags are used only by the working-class adolescents.

(16)   a.   Charlie: the only time I drink is like at parties or =

          b.   Matt: = yeah..

                             not one of the things you do every day really is it..daft

          c.   Charlie: don't like smoking or anything like that ..no that's disgusting

→ d.   Matt: I used to *me*..well I tried it

→ e.   Charlie: I haven't even tried it *me*

          f.   Matt: my mam wouldn't say nowt

          g.   AW: do your parents smoke?

          h.   Charlie: my mam does

          i.   Matt: all of them do..got my real dad my step dad and my mam

→ j.   Charlie: I don't like it *me*

(17)   a.   AW: right what about a favourite singer then?

→ b.   Kay: Peter André *me*

          c.   Ruth: Peter André's allright but

→ d.   Kay: he's got a real nice chest *him*

          e.   AW: has he? is it hairy?

          f.   Kay: no it's real brown and greasy

          g.   Ruth: cos he has baby oil smothered on him

In (16d), (16e), (16j) and (17d) the tags are co-referential with the subject pronoun in the preceding clause. As such they are subject to aspects of binding theory. Like other discourse-related phenomena considered as dislocation (for example, NP-fronting, *it*-clefts and left dislocation), they can be accommodated in generative syntax within the left periphery of the clause (Henry 1995: 135).

Their occurrence in spoken English in Ayrshire has been analysed by Macaulay (1989, 1991), who found them to be a feature of working-class speech. Macaulay reports that the working-class speakers in his study used constructions that brought personal pronouns into prominence in a way not

found in the speech of the middle classes. These constructions included not only pronoun tags but also left dislocation, NP-fronting and *it*-clefts, all of which were used more frequently by the lower class speakers with the function of expressing intensity. The middle-class speakers, by contrast, tended to convey intensity through the use of adverbials. This previously unsuspected finding has interesting sociolinguistic implications, but it shows the difficulty of using the linguistic variable for the analysis of forms expressing intensity (and Macaulay did not attempt to do so). The variable may well be a heuristic construct that does not necessarily map directly onto the units of linguistic structure (Wolfram 1993), but to include left dislocation, say, or *it*-clefts in the same analytic unit as adverbials would be stretching the concept of the variable beyond all credibility, even if the forms can be considered to have the same discourse function.

Even if we ignore the sociolinguistic differences in the expression of intensity and focus simply on the tag constructions, it is still difficult to analyse the tags within a variationist framework. Intensity (which seems to be used in the literature with much the same meaning as 'emphasis') is often said to be the discourse function of a form that appears to be involved in syntactic variation. Emphasis does not affect truth conditions, so if the pronoun tags in our data are emphatic in function they could in principle be analysed within a variationist framework: the clauses with tags could be seen as semantically equivalent variants to the corresponding clauses without tags. However, it would be difficult to identify the envelope of variation. Emphasis is an ill-defined concept, lacking theoretical rigour. It does not provide a basis for predicting which variant will be preferred on any one occasion, nor even where it is possible for variation to occur (Schwenter 2003). To say that emphasis highlights a particular entity in the discourse – in this case, the co-referential subject pronoun – is too subjective to be useful if we are looking for language universals, as Myhill (1992: 3) points out. Sells et al. (1996: 174) claim that unless the status of emphasis can be clearly specified in the grammar, along with the extent to which it can affect the form and function of different linguistic phenomena, the very ubiquity of appeals to this type of affective meaning may reduce its analytic value.

Nevertheless speakers do appear to choose a range of expressions to add some kind of additional, affective meaning to their propositions. We are omitting an aspect of language that is important to speakers if we do not take the expression of emphasis into consideration when attempting to explain syntactic variation, whether within a generative or a variationist framework.

A further factor constraining a speaker's use of pronoun tags is their interactional function. This too, however, can be difficult to take into account in a rigorous way. For example, in the Hull data the tags sometimes occur with a form referring to an entity that becomes a conversational topic: in (17b), for example, Kay's *Peter André* is picked up by Ruth in the next turn. Similarly, in (17e) Ann Williams picks up the topic of Peter André's nice chest, proposed by Kay in the preceding turn. In both these examples, then, the tag has a function in conversational management. Note that in (17d) the fact that the tag is co-referential with the subject pronoun is irrelevant in terms of its interactional function: it is the entire proposition (Peter André and his chest) that is proposed as a topic.

Elsewhere in the data the tags sometimes appear to explicitly mark a contrast, often between the content of the utterance in which they occur and the content of the previous turn. This is possibly the case in (16d), where Matt and Charlie are discussing smoking with Ann Williams (AW). Charlie is a keen anti-smoker and his first utterance about smoking, in (16c), makes it clear that he does not like it (*don't like smoking or anything like that... no that's disgusting...*). Matt, with his *I used to me* in (16d) perhaps shows that he has inferred from this that Charlie does not smoke – or perhaps he already knows this to be true – and he claims, in contrast, to have smoked himself. Charlie's *I haven't even tried it me*, in (16e), then clearly contrasts his own lack of experience with Matt's and in (16j) he repeats the assertion he made in (16c), this time contrasting his own dislike of smoking with the behaviour of Matt's family, which has been described in the preceding two turns.

The tags may simultaneously (or alternatively) signal a contrast between what the speaker would find it interesting to talk about and what the previous speaker has been saying; thus from (16c) onwards both Matt's and Charlie's contributions concern their own experiences of smoking and their family's behaviour concerning smoking. The tags may signal a desired change of topic, then, often in contrast with a previous topic; but there is no principled way of deciding on the basis for the contrast. In Charlie's turns he also expresses his stance towards the content of his clause; the tags may draw attention to his stance and contrast it with Matt's. There are many interactional factors, then, that can affect a speaker's choice of construction and that may constrain variation between clauses with tags and clauses without tags.

Many other constructions are similarly conditioned by interactional factors. A further example from English is the *get*-passive construction, illustrated in (18):

(18)   Josephine got run over by a bus

This can be seen as an alternant to a conventional *be*-passive, as in (19), and both (18) and (19) as alternants to the corresponding active clauses in (20) and (21):

(19)   Josephine was run over by a bus

(20)   A bus ran Josephine over

(21)   A bus ran over Josephine

Weiner and Labov (1983:43) claimed that a shift to the *get*-passive is one of the most active grammatical changes taking place in present-day English. An important question for the study of language variation and change therefore concerns the distribution of the form throughout the speech community, since this will allow us to chart the diffusion of the change. As with pronoun tags and lone *when* phrases, there is no initial problem in using the variable to analyse this alternation: the alternants in (18)–(21) have the same main verb and the same entities as arguments, so we can assume that they are semantically equivalent.

Passive constructions are relatively infrequent in speech. It is not surprising, therefore, that there have been few quantitative studies of the English passive. Macaulay (1991), however, analysed passive constructions in middle class and working-class speech in his Ayrshire study. Although there were no significant social class differences in the overall use of passive clauses, *get*-passives were used more frequently by the working-class speakers. *Get*-passives have been a major shibboleth in British schools so it is possible that they were simply avoided by the middle class speakers in the context of the sociolinguistic interview: the working-class speakers in the study had far less exposure to formal schooling. Interestingly, however, Macaulay further reports that the *get*-passive occurred almost exclusively with animate subjects and that these, in turn, were also more frequent in the working-class interviews. *Get*-passives are eventive aspectually and this presumably contributes to the animacy effect; events are usually controlled by an actor and animates are more likely to be able to control events. One factor affecting the use of the *get*-passive, then, is, quite simply, what speakers choose to talk about. If they talk about animate beings, there is a greater chance of their using the *get*-passive.

Carter and McCarthy (1999) add a further dimension to our understanding of the use of this form: their corpus-based analysis reveals that the *get*-passive highlights the stance of speakers towards the grammatical subject and the event encoded in the verb phrase – a stance that usually indicates their judgement that the circumstances are adverse, problematic or otherwise note-

worthy. We cannot necessarily assume that the Ayrshire speakers and the speakers in Carter and McCarthy's study are affected equally by these factors, but the two studies at least suggest the possibility that the shift to the *get*-passive in present-day English is led by a group-specific discourse preference for talking about animate beings and encoding syntactically the speaker's stance towards actors and the event referred to by the verb. Is there, in other words, a distinctive habitual pattern of interaction for the working-class group of speakers that is not shared by the middle class group? Unfortunately, however, we now run into the same problems as with appeals to emphasis: although a wide range of syntactic features have been said to express the speaker's stance towards the content of their utterances (see, for example, those listed in Ochs & Schieffelin 1989) it is difficult to give a rigorous definition of the concept and it has no status within any syntactic theory of language.

Syntactic variation and syntactic change seem here to be intimately and inextricably part of the social construction of discourse. Carter and McCarthy's findings point to a similar conclusion: as these authors say (op. cit.: 55), judgements about adversity, noteworthiness and the like are socio-culturally founded and emergent in the interaction rather than inherent in the semantics of verb choice or the selection of voice or aspect. In order to address these issues, then, and to fully analyse patterns of variation and change in the use of English passive constructions, it is necessary to add a qualitative, interpretive dimension to the analysis.

All syntactic variables in fact, including low level variables such as verbal *-s*, call for this kind of interpretive discourse analysis. Levinson (1988: 166) made this point in relation to English *ain't*, asking whether working-class speakers who use *ain't* frequently do so because for them *ain't* is a marker of group identity, or because it is a more emphatic form of negation than *isn't*, *aren't, hasn't* and *haven't.* If the latter, does this reflect the habitual patterns of social interaction of the social group to which the heavy *ain't* users belong? They might, for example, utter emphatic denials more frequently than other social groups in the community because they more often receive accusations. Thus in order to understand how and why speakers use variation and the effect that their usage has on language change, we cannot simply analyse the simple alternation of forms: we must also perform qualitative analyses to see how these forms are used in social interaction, and find a way of incorporating this dimension into an explanatory theory of language structure and language change.

6.   Conclusion

My starting point in this paper was that the syntactic structure of spoken language has not been seriously taken into account by either variationists or generativists. A hope, from my sociolinguist corner, is that a new alliance between variationists and generativists might lead to generative linguists becoming more interested in externalised, performed language – not simply in order to find new data against which to test the theory but also to apply the rigour of the generative approach to discovering the structure of spoken language. There are many ways in which our understanding of spoken language might benefit. First, it would be helpful if some theoretical rigour could be brought to the concepts of emphasis (or intensity) and stance, which seem so important to speakers and which are so often appealed to in descriptions of language use. Second, as pointed out by Henry (2002: 277) generativists might help variationists determine on a more systematic basis those syntactic structures that should be considered variants of a single form. This would help assess the universal dimension of any internal constraints on variation (Cornips & Corrigan in press) and may prevent us from being unduly influenced by the standard ideology when deciding on the forms to be analysed as variants of a single variable. It might even help resolve the issue once and for all of whether the variable should be used to analyse variation in syntax. Third, it may be possible to identify some language universals governing the form of spoken language – though for this to be possible we would need, in my view, to broaden our perspective and look for social or interactional principles in addition to the principles governing innate structures. Perhaps in this endeavour we will need to work with researchers from other fields of linguistics. For example, the perspective of language typology has been very successfully applied to the findings of social dialectology (for example, Kortman 2002, 2004).[3]

   Generativists and variationists might work together to develop a methodology that could determine when the phrases of spoken language have been generated by the grammar and when they are prefabricated sequences produced from memory. A decision needs to be taken about whether constructions that appear to be semi-flexible and that are capable of description within 'mini-grammars' are best accounted for as generated by the internal grammar or as part of the lexicon (or, as suggested earlier, at an interface level). A related problem is that of establishing, for any given syntactic feature that seems to be variable, how much of the variation is most usefully considered as syntactic (in other words, internally motivated by the grammar) and how much is better seen as pragmatically determined.

Many of these issues will require both a qualitative and a quantitative dimension to the analysis. I have argued in this paper for a complementary approach to the analysis of syntactic variation, which analyses syntactic forms in their discourse context. If we are to gain insights from our different research traditions we need to be aware that the way speakers use the forms of spoken language may result as much from interactional and social factors as from biological factors, and come to a principled decision about whether such forms should be seen as generated by the grammar or attributed to performance mechanisms. In this way we may at last succeed in understanding how the cognitive and the social aspects of language are integrated as part of the human experience.

Notes

* I would like to warmly thank David Adger, Miriam Meyerhoff, Leslie Milroy, Leonie Cornips and Karen Corrigan for comments on this paper, as well as the anonymous referees. All shortcomings are my own responsibility.

1. In fact analyses of syntactic variation have tended to give more weight to internal constraints than social constraints, perhaps because of the problems inherent in using the variable to analyse syntactic variation; see Milroy and Gordon (2003).

2. In all these expressions *when* appears to have indefinite meaning (Lyons 1977; Haslpelmath 1997); it resembles its use in conversation as a deictic that is non-specific for one of the participants (as in, for example, *tell me about the time when you were lost in the mountains*, meaning 'the time you were lost', or *I tell you what .. I'll wash the dishes, if you do the cooking*).

3. The typological approach is of course divergent from and perhaps even runs counter to the aims of the present volume.

Transcription conventions

..	short pause (not timed)
=	utterance latched on to previous turn (with no discernible pause)
?	question marks show the end of a stretch of talk interpreted as a question
<LAUGHTER>	angled brackets give additional information
[	extended square brackets show the beginning of an overlap
→	an arrow indicates that the line to the right is the one where a given example occurs

References

Aijmer, Karin (1996). *Conversational Routines in English: Convention and Creativity*. Harlow: Longman.

Biber, Douglas, Susan Conrad, & Randi Reppen (1994). "Corpus-Based Approaches to Issues in Applied Linguistics." *Applied Linguistics, 15*, 169–187.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, & Edward Finegan (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Ltd.

Blanche-Benveniste, Claude (1997). "La Notion de Variation Syntactique dans la Langue Parlée." *Langue Française, 115*, 19–29.

Bolinger, Dwight (1975). "Meaning and Memory." *Forum Linguisticum, 1*, 2–14.

Brewer, Jerome P. (1986). "Durative Marker or Hypercorrection? The Case of *-s* in the WPA Ex-Slave Narratives." In Michael B. Montgomery & Guy Bailey (Eds.), *Language Variety in the South* (pp. 131–148). Tucson: The University of Alabama Press.

Brown, Gillian & George Yule (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.

Carter, Ronald & Michael McCarthy (1999). "The English *Get*-Passive in Spoken Discourse: Description and Implications for an Interpersonal Grammar." *English Language and Linguistics, 3*, 41–58.

Chambers, J. K. & Peter Trudgill (1998). *Dialectology* (2nd ed.). Cambridge: Cambridge University Press.

Cheshire, Jenny (1982). *Variation in an English Dialect: A Sociolinguistic Study*. Cambridge: Cambridge University Press.

Cheshire, Jenny (1987). "Syntactic Variation, the Linguistic Variable and Sociolinguistic Theory." *Linguistics, 25*, 257–282.

Cheshire, Jenny (1999). "Spoken Standard English." In T. Bex & R. J. Watts (Eds.), *Standard English: The Widening Debate* (pp. 129–148). London: Routledge.

Cheshire, Jenny (2000). "The Telling or the Tale? Narratives and Gender in Adolescent Friendship Networks." *Journal of Sociolinguistics, 4*, 234–262.

Cheshire, Jenny, Paul Kerswill, & Ann Williams (1999). *The Role of Adolescents in Dialect Levelling*. Final Report submitted to the Economic and Social Research Council, London, June 1999. Ref. R000236180.

Cheshire, Jenny, Paul Kerswill, & Ann Williams (2005). "Phonology, Grammar and Discourse in Dialect Convergence". In Peter Auer, Franz Hinskens, & Paul Kerswill (Eds.), *Dialect Change: Convergence and Divergence in European Languages* (pp. 135–167). Cambridge: Cambridge University Press.

Cheshire, Jenny & Dieter Stein (1997). "The Syntax of Spoken Language". In Jenny Cheshire & Dieter Stein (Eds.), *Taming the Vernacular: From Dialect to Written Standard Language* (pp. 1–12). Harlow: Longman.

Chomsky, Noam (1999). "Derivations by Phase". Unpublished MA. Boston: MIT.

Clarke, Sandra (1997). "English Verbal *-s* Revisited: The Evidence from Newfoundland". *American Speech, 72*, 227–259.

Cornips, Leonie & Karen Corrigan (2005). "Convergence and Divergence in Grammar." In P. Auer, F. Hinskens, & P. Kerswill (Eds.), *Dialect Change: Convergence and Divergence in European Languages* (pp. 96–134). Cambridge: Cambridge University Press.

Corrigan, Karen (1997). The Syntax of South Armagh English in its Socio-Historical Perspective. Unpublished doctoral dissertation, University College Dublin.

Couper-Kuhlen, Elizabeth & Margret Selting (Eds.). (2001). *Studies in Interactional Linguistics*. Amsterdam: Benjamins.

Coupland, Nikolas (1983). "Patterns of Encounter Management: Further Arguments for Discourse Variables." *Language in Society, 12*, 459–476.

Coveney, Aiden (1997). "L'Approche Variationniste et la Description de la Grammaire du Français." *Langue Française, 115*, 88–100.

Coveney, Aiden (2002). "Supplement: A Critical Survey of Recent Research on Grammatical Variability, Negation and Interrogation in French." *Variability in Spoken French*. Bristol: Intellect.

Crystal, David (1976). "Neglected Grammatical Factors in Conversational English." In S. Greenbaum, G. Leech, & J. Svartvik (Eds.), *Studies in English Linguistics* (pp. 153–166). London: Longman.

Crystal, David (1995). *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press.

Dines, Elizabeth R. (1980). "Variation in Discourse – "And Stuff Like That"". *Language in Society, 9*, 13–33.

Edwards, John & Derek Middleton (1986). "Joint Remembering: Constructing an Account of Shared Experience Through Conversational Discourse." *Discourse Processes, 9*, 423–459.

Ford, Cecilia E. (1993). *Grammar in Interaction: Adverbial Clauses in American English Conversations*. Cambridge: Cambridge University Press.

Fox Tree, Jean E. & Josef C. Schrock (2002). "Basic Meanings of *You Know* and *I Mean*". *Journal of Pragmatics, 34*, 727–747.

Gadet, Françoise (1997). "La Variation, Plus Qu'une Écume." *Langue Française, 115*, 5–18.

Godfrey, Elizabeth & Sali Tagliamonte (1999). "Another Piece for the Verbal *-s* Story: Evidence from Devon in South-West England". *Language Variation and Change, 11*, 87–121.

Haspelmath, Martin (1997). *Indefinite Pronouns*. Oxford: Oxford University Press.

Henry, Alison (1995). *Belfast English and Standard English: Dialect Variation and Parameter Setting*. New York: Oxford University Press.

Henry, Alison (2002). "Variation and Syntactic Theory." In J. K. Chambers, Peter Trudgill, & Natalie Schilling-Estes (Eds.), *Handbook of Language Variation and Change* (pp. 267–282). Oxford: Blackwell.

Holmes, Janet (1986). "Functions of *You Know* in Women's and Men's Speech". *Language in Society, 15*, 1–21.

Holmes, Janet (1995). *Women, Men and Language*. Harlow: Longman.

Kortmann, Bendt (2002). "New Prospects for the Study of Dialect Syntax: Impetus from Syntactic Theory and Language Typology." In S. Barbiers, L. Cornips, & S. van der Kleij (Eds.), *Syntactic Microvariation*. Electronic publication of the Meertens Institute, http://www.meertens.knaw.nl/projecten/sand/synmic/. Amsterdam: Meertens Institute.

Kortmann, Bendt (Ed.). (2004). *Dialectology meets Typology: Dialect Grammar from a Cross-linguistic Perspective*. Berlin/New York: Mouton de Gruyter.

Labov, William (1970). "The Study of Language in its Social Context." *Studium Generale, 23*, 30–87.

Lavandera, Beatriz (1978). "Where Does the Sociolinguistic Variable Stop?" *Language in Society, 7*, 171–182.

Leech, Geoffrey (2000). "Grammars of Spoken English: The New Implications of Corpus-Oriented Research." *Language Learning, 50*, 675–724.

Levinson, Stephen (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, Stephen (1988). "Conceptual Problems in the Study of Regional and Cultural Style." In Norbert Ditttmar & Peter Schlobinski (Eds.), *The Sociolinguistics of Urban Vernaculars* (pp. 161–190). Berlin: Walter de Gruyter.

Lyons, John (1977). *Semantics. Volume II*. Cambridge: Cambridge University Press.

Macaulay, Ronald K. S. (1989). "'He Was Some Man Him': Emphatic Pronouns in Scottish English." In Thomas J. Walsh (Ed.), *Synchronic and Diachronic Approaches to Linguistic Variation and Change* (pp. 179–187). Washington, DC: Georgetown University Press.

Macaulay, Ronald K. S. (1991). *Locating Dialect in Discourse: The Language of Honest Men and Bonnie Lassies in Ayr*. New York: Oxford University Press.

McCarthy, Michael (1991). *Discourse Analysis for Language Teachers*. Cambridge: Cambridge University Press.

McCarthy, Michael (1998). *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.

Miller, Jim (in press). "Spoken and Written English." In B. Aarts & A. McMahon (Eds.), *The Handbook of English Linguistics*. Oxford: Blackwell.

Miller, Jim & Regina Weinert (1998). *Spontaneous Spoken Language: Syntax and Discourse*. Oxford: Clarendon Press.

Milroy, Lesley (1999). "Standard English and Language Ideology in Britain and the United States." In T. Bex & R. J. Watts (Eds.), *Standard English: The Widening Debate* (pp. 173–206). London: Routledge.

Milroy, Lesley (2001). "Conversation, Spoken Language and Social Identity." In P. Eckert & J. R. Rickford (Eds.), *Style and Sociolinguistic Variation* (pp. 268–278). Cambridge: Cambridge University Press.

Milroy, Lesley & Matthew Gordon (2003). *Sociolinguistics: Method and Interpretation*. Oxford: Blackwell.

Mondorf, Britta (2000). Sex Differences in English Syntax. Unpublished doctoral dissertation, Düsseldorf University.

Montgomery, Michael, Janet M. Fuller, & Sharon DeMarse (1993). "'The Black Men Has Wives and Sweet Harts [and Third Person Plural *-s*] Jest Like the White Men': Evidence for Verbal *-s* from Written Documents on 19th Century African American Speech." *Language Variation and Change, 5*, 335–357.

Myhill, Jim (1992). *Typological Discourse Analysis*. Oxford: Blackwell.

Myhill, Jim & Wendell A. Harris (1986). "The Use of the Verbal *-s* Inflection in BEV." In David Sankoff (Ed.), *Diversity and Diachrony* (pp. 25–32). Amsterdam/Philadelphia: Benjamins.

Ochs, Elinor, Emmanuel A. Schegloff, & Sandra A. Thompson (Eds.). (1996). *Interaction and Grammar*. Cambridge: Cambridge University Press.

Ochs, Elinor & Bambi Schieffelin (1989). "Language Has a Heart". *Text, 9*, 7–25.

Pitts, Ann (1986). "Contrastive Use of Verbal *-s* in Slave Narratives." In David Sankoff (Ed.), *Diversity and Diachrony* (pp. 73–82). Amsterdam/Philadelphia: Benjamins.

Poplack, Shana & Sali Tagliamonte (1989). "There's No Tense Like the Present: Verbal *-s* Inflection in Early Black English." *Language Variation and Change, 1*, 47–84.

Quirk, Randolph, Sydney Greenbaum, Geoffrey Leech, & Jan Svartvik (1985). *A Grammar of Contemporary English*. Harlow: Longman.

Romaine, Suzanne (1980). "The Relative Clause Marker in Scots English: Diffusion, Complexity and Style as Dimensions of Syntactic Change." *Language in Society, 9*, 221–247.

Romaine, Suzanne (1984). "On the Problem of Syntactic Variation and Pragmatic Meaning in Sociolinguistic Theory." *Folia Linguistica, 18*, 409–439.

Schegloff, Emmanuel (1989). "Reflections on Language, Development and the Interactional Character of Talk-In-Interaction." In Marc H. Bornstein & Jerome S. Bruner (Eds.), *Interaction in Human Development* (pp. 139–153). Hillsdale, NJ: Lawrence Erlbaum.

Schneider, Edgar (1983). "The Origin of the Verbal *-s* in Black English." *American Speech, 58*, 99–113.

Schneider, Edgar (2002). "Investigating Variation and Change in Written Documents." In J. K. Chambers, Peter Trudgill, & Natalie Schilling-Estes (Eds.), *Handbook of Language Variation and Change* (pp. 67–96). Oxford: Blackwell.

Schwenter, Scott (2003). "*No* and *Tampoco*: A Pragmatic Distinction in Spanish Negation." *Journal of Pragmatics, 35*, 999–1030.

Sells, Peter, John Rickford, & Thomas Wasow (1996). "Variation in Negative Inversion in AAVE: An Optimality Theoretic Approach." In Jennifer Arnold et al. (Eds.), *Sociolinguistic Variation: Data, Theory and Analysis. Selected Papers from NWAVE 23* (pp. 171–1760). Stanford: CSLI.

Sinclair, John (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Skehan, Peter (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Trudgill, Peter (1974). *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Weiner, Judith & William Labov (1983). "Constraints on the Agentless Passive." *Journal of Linguistics, 19*, 29–58.

Winford, Donald (1996). "The Problem of Syntactic Variation." In J. Arnold et al. (Eds.), *Sociolinguistic Variation: Data, Theory and Analysis. Selected Papers from NWAVE 23* (pp. 177–192). Stanford: CSLI.

Wolfram, Walt (1993). "Identifying and Interpreting Variables." In Dennis R. Preston (Ed.), *American Dialect Research* (pp. 193–222). Amsterdam/Philadelphia: John Benjamins.

Wray, Alison (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

PART II

Individual speaker variability and methodological innovation

Idiolectal variation and syntactic theory

Alison Henry

University of Ulster at Jordanstown

1. Introduction

The early years of work on generative grammar (see, for example, Chomsky 1957, 1965) focussed on identifying the possible types of linguistic rule; later work, however, has moved towards seeing language as a system of (perhaps optimal) principles allowing a very limited range of cross-linguistic variation. While early work interfaced more easily with studies of sociolinguistic variation, giving rise, for example, to the concept of 'variable rules' (Labov 1972: 93– 95), later work (Chomsky 1995, 2000, 2001a, b) – with its strong universalist focus – sits more uneasily with the type of variation found by sociolinguists.

In recent years there has however also been a move by syntacticians towards studying variation in more detail, in particular, at the dialectal rather than language level. On the whole this has shown that microvariation at the level of the dialect can be encompassed in the Principles and Parameters/Minimalist system (for extended treatments of dialects within this framework, see, for example, Haegeman 1992; Henry 1995; Tortora 2003). However, work on variation between the idiolects of individuals has been largely lacking. This is rather surprising, given that the goal of syntactic theory is to account for the possible internal grammars of native speakers: it considers that this internal or I-language is the proper subject of study, and its goal is to determine the initial and steady states of the language faculty in the mind/brain of individuals. In practice, however, studies of and papers published about grammar have in general been descriptions of aspects of a 'language' or 'dialect', even though these are E-language concepts which are not recognized by the theory. More-over, the actual data gathering practice – which generally includes consulting a group of speakers about their intuitions and taking the most frequent view,

or the view of a subset – has led to research practice where variation among individual speakers is, in effect, screened out.

This paper considers variation across individual speakers of English in relation to a particular aspect of English syntax which has been central in the development of syntactic theory – subject-verb agreement in sentences with expletive *there* – and it is shown that individual idiolects vary considerably in relation to the conditions under which agreement does or does not take place. The types of grammars arrived at and the factors involved not only seem to show that the highly constrained view of grammar pursued in syntactic theory recently is, in fact, overly constrained, and that a return to a concept of rules may be required; they also show a new means of accommodating linguistic and sociolinguistic theory.

2.   Methodology

Discussions on methodology do not, in general, figure strongly in work on generative syntax. However, it is clear that the method of data collection can affect the data collected – for example, if it involves the possibility of ignoring some of the data in favour of writing a grammar for only part of it.

In sociolinguistics, the type of data collected for variationist studies is generally naturalistic, production data, whereas the data collected for generativist studies relies more on the intuitions of native speakers. The need for the latter in relation to studies of syntax or morphology arises mainly because any given syntactic structure is likely to be very infrequent in data. And, if one is seeking to identify the nature of individual grammars, it would require enormous amounts of data from different individuals to establish what forms they used. Even then, one would not know if forms they did not use were absent from their grammars, or simply chance omissions from the data.

2.1 Methodological practices adhered to in this study

In this study, then, intuition data was used. But in order to ensure that variation among idiolects was not suppressed, each speaker's intuitions were studied individually, rather than in a group setting. The data collection method used was that discussed in Henry (in press): this involves working together with native speakers to establish what their grammars are. A fixed set of sentences is not tested, other than in the very initial stages of investigation: rather, the researcher works with the native speaker to establish what characteristics of

sentences are influencing their use or non-use of the phenomenon under study. Thus, from an initial small set of sentences, further sentences were created in order to establish where agreement was and was not possible, and speakers were encouraged to reflect on where agreement could occur and where it could not. As such, patterns which emerged from the initial data could be tested, and it could be established what in particular was constraining the use of agreement for the speaker concerned. In Henry (in press), I argue that such a method is not an unscientific way of gathering data; rather, it enables the intuitions of native speakers about their languages to be tapped and their internalised knowledge of their own language variety to be accessed as far as possible. It means that where speakers reject sentences, the reason for doing so can be explored. Simply knowing that a speaker rejects a sentence is not enough – we need to know what characteristic of the sentence makes it ungrammatical. Thus, for example, it is not sufficient to know that a speaker rejects a sentence like:

(1)   There was ninety books on the shelf

We need to be able to explore precisely why the speaker rejects this sentence, whether it is because of the tense of the verb, or something about the nature of the noun phrase following the verb, and we need to be able to make small changes to these, and test a range of similar sentences, to establish what criteria come into play for that particular speaker in determining whether a sentence lacking agreement with the associate is possible. This cannot be done with a predetermined list of sentences; and often speakers will volunteer information about what they can and cannot say, or what they would say instead, that enables the linguist to arrive at a more complete an understanding of their grammar than would otherwise be possible.

The data in this paper is based on such a study of native speakers. In all, 25 speakers were studied, and they revealed a range of different grammars, which are discussed below. All the speakers were speakers of Northern Irish-English who did not use 'singular concord' (Henry 1995: Ch. 2), that is, who could not use sentences like the following, where a plural subject is found with an apparently singular verb form.

(2)   The eggs is cracked
(3)   The doors was opened

3.   Patterns of idiolectal variation in agreement structures

3.1 Agreement in sentences with expletive *there*

In Standard English, in sentences with expletive *there* such as (4) and (5), the verb agrees with the associate.[1]

    (4)   There are three books on the table

    (5)   There were several magazines on the shelf

Accounting for this has been one of the factors that has generated considerable discussion in the literature of generative grammar, and has indeed been responsible for some of the major aspects of development of minimalist syntactic theory (see, for example, Chomsky 1995:Ch. 4).

    There is, however, acknowledged to be considerable variation in agreement patterns in sentences with expletive *there*, variation which consistently shows up in sociolinguistic studies (see, for example, Meechan & Foley 1994).

    A purely minimalist approach might lead us to expect that there would be two possible grammars: one in which *there* does not fully check or value the properties of the verb, leaving the number feature to be determined by the associate as in Standard English, and one in which *there* has a full set of features, thus meaning that the verb, in effect, agrees with *there* and is always singular, regardless of the number of the associate. In other words, we would expect that one group of speakers would find a sentence like (6) grammatical, and one like (7) ungrammatical, whereas, for the other group, the opposite would be the case, with (7) grammatical and (6) ungrammatical.

    (6)   There are three books on the table

    (7)   There's three books on the table

It is often claimed within generative grammar that no optionality exists (Chomsky 1995). However, it has been suggested by some that there is a possibility of having competing grammars (Kroch 1989); or one might envisage speakers being bidialectal and having a grammar for standard, formal usage and one for informal usage, and on this basis one might explain finding speakers for whom both (6) and (7) are grammatical. However, it would not be predicted that there would be a range of possible grammars, or that the nature of the associate or the tense of the verb would have any effect on the occurrence or otherwise of agreement.

In fact, however, the present study found a range of grammars even among the relatively small group of speakers studied. For some, as Chomsky (1995: Ch. 4) notes, there is agreement everywhere except where *there's* is used and this must be in the contracted form rather than the full *there is* type.

(8)   There's three books on the table

(9)   *There is three books on the table

(10)   *There was three books on the table

A second group of speakers allows both the uncontracted and contracted forms, *there's* and *there is* (so that both (8) and (9) are grammatical) but not *there was* with plural associates (10), while another group allows the singular form to occur only in the past with *was* (meaning that, for them only (10) in the above set is grammatical). Some speakers allow both contracted forms as in (8) and past forms as in (10), but not forms like (9), with uncontracted *is.* There are also differences in relation to whether the sentences are negatives, or questions, and with regard to the nature of the associate. For some speakers associates which are not singular can only have singular verbs if they are dual, or represent less than about 10 items; for others, only associates which are in some sense 'identifiable groups' allow non-agreement; for yet others only structures with designations of groups such as 'lots of NP' allow the singular verb. Moreover, individual speakers show a high degree of optionality, often having both agreement and non-agreement in a range of contexts.

The individual grammars found have a combination of the following properties acting on the possibility of *there* agreeing with the associate: verb factors, associate factors and other factors in the sentence, such as whether or not it is negative. These will be considered in turn below.

3.1.1 *Verb factors*

Provided that the verb is contracted, most speakers who accepted expletive sentences without agreement accept sentences with *there's* such as:

(11)   There's some books on the table

(12)   There's trees in the garden

Only some of these speakers also accept the uncontracted form *there is* with a plural associate.

Thus, whether or not the verb is contracted seems to be an important factor in determining whether the verb agrees with the subject or not. This may conceivably not be because of the contraction per se, but because such con-

traction is a particular characteristic of informal spoken language, and occurs infrequently in writing, for example. For speakers working with two possible grammars, a 'formal' and 'informal' one, it could be the case that the informality of contraction coincides with a grammar that allows the verb not to agree with the associate, whereas the more formal, uncontracted form *there is/are* requires agreement.

There are some speakers who do not allow non-agreement when the verb is present (except where it is contracted as in (12) above), but only in the past. For these speakers (14) is grammatical but (13) is not:

(13) *There is three books on the table

(14) There was three books on the table

On the contrary, two speakers had exactly the opposite intuitions: sentences like (13) were fine for them, but sentences like (14), where the verb *be* is in the past, required agreement and, where this was not present, were deemed to be ungrammatical as in (15). These speakers did not generally use *was* with plural subjects, hence the ungrammaticality for them also of (16) (which is of course grammatical in some English dialects as Adger & Smith this volume, also contend), so we cannot attribute their allowing lack of agreement with *was* to that factor.

(15) *Three books was on the table

(16) *The students was happy

Verbs other than *be,* such as the auxiliary *have* or raising verbs like *seem* and *appear,* can be the agreeing elements in expletives. Some speakers allow lack of agreement with raising verbs ((17) and (18)), while not allowing it in any other case except with contracted *there's.*

(17) There seems to be three men in the room

(18) There appears to be some books on the table

For one speaker, however, this only applied to *seem*; for this speaker, (17) was grammatical but (18) was not.

Thus, a number of factors concerning the verb, which is the potential locus of agreement, affect whether agreement occurs and would seem to include: contraction, tense, and which lexical verb is used.

3.1.2 *Associate factors*

According to the analysis under which agreement or otherwise with the associate is determined by the nature of the expletive *there*, that is, whether or not it fully values the features of T, it would not be expected that the nature of the associate would have any influence on agreement. However, for some speakers, this is clearly the case: the following factors affect agreement possibilities for a number of speakers.

For some speakers, non-agreement is possible when the associate contains a term like *lots,* even where the speaker does not generally allow agreement with the verb form used. The following intuitions are from a speaker who generally had obligatory agreement with the associate when the verb was past like (21), but allowed lack of agreement where the associate was introduced by *lots* as in (19).

(19)   There was lots of books on the shelf

(20)   There were lots of books on the shelf

(21)   *There was books on the shelf

(22)   There were books on the shelf

This could be thought, like the facilitating of non-agreement by the contracted form *there's*, to be due to the informal, colloquial status of *lots*, which is generally replaced in writing and more formal spoken contexts by *many* as in (23):

(23)   There were many books on the shelf

However, speakers who accepted sentences with *lots* and without agreement also accepted lack of agreement with more formal terms such as *scores:*

(24)   There was scores of books on the shelf

(25)   There was scores of students certain to fail the exam

For some speakers, it was not so much the presence of a term like *lots*, but rather the length of the associate which affected agreement, or more specifically, whether or not the associate-initial element was the head noun. These speakers did not find sentences such as (26) grammatical, but were happy with (27):

(26)   *There's books on the table

(27)   There's three books on the table

For some of them, a number or quantifier was required before the head noun in order to have non-agreement; for others, any element appeared to suffice in this position, so that (28) and (29) were both grammatical:

(28)   There's large books on the table

(29)   There's linguistics books on the table

We have seen that for some speakers, elements other than the head noun in the associate affect the use of agreement; for some, it is specifically the number of elements denoted by the associate that matters: in particular, a low number of items as in (30) (for some speakers this is two, for others less than about 10) allows non-agreement, whereas larger numbers as in (31) require agreement:

(30)   There's three books on the shelf

(31)   *There's seventy-five books on the shelf

Note that this seems to be in contradiction to the occurrence of non-agreement with items like *lots of* which designate large amounts; there are, however, speakers who allow non-agreement with both *lots of* and with small amounts, but not with large amounts, designated by a numeral.

It might be thought that agreement here is determined by the nature of the associate: perhaps certain Noun Phrases do not trigger agreement. However, if that were the case one would predict that the same agreement pattern would be found when the same elements were in subject position: but that is not the case either; it is to be noted that the speakers under study did not allow non-agreement where the equivalent noun phrases were in subject position, namely:

(32)   *Lots of books is on the shelf

(33)   *Three books is on the shelf

There is thus a complexity of factors relating to the associate, which differ across speakers, predicting whether or not agreement takes place.

3.1.3 *Other factors*

For some speakers, inversion, as found in yes-no questions, favoured agreement:

(34)   There was three books on the table

(35)   *Was there three books on the table?

Others always had agreement in negatives as in (36)–(37a) while there were other speakers who generally required agreement, but could have lack of agreement if there was a negative, as in (37b):

(36)   There weren't any books on the table

(37)   a.   *There wasn't any books on the table
        b.   There wasn't any books on the table

In other words, although the grammars of speakers referred to similar categories, that is, negation, in some cases this disfavoured agreement and in others favoured it.

3.1.4 *Interaction of factors*

There does not seem to be any particular constraint on interaction of verb factors, associate factors, and other factors; for different speakers, a different variety of factors of each type can co-occur. This was a study of a relatively small number (25) of native speakers of English. A study of a larger number of speakers is likely to find yet other factors determining whether or not agreement is obligatory; it seems improbable that the factors found among these 25 speakers are exhaustive. But even as outlined here, it seems unlikely that these can be encompassed in the highly restricted framework of current minimalist syntax.

The kind of grammar we need to encompass any of the types of grammar found, seems to need to be of the following type: it needs to be able to specify the conditions under which agreement is or is not obligatory; thus, the agreement of a constituent X with a constituent Y must be able to be specified as subject to certain conditions:

(38)   a.   Agree (X with Y) if, Y precedes X OR X -past, X contracted, Y 10+, Y lots+N, Y inanimate.
        b.   The sentence including X and Y is: + negative, + interrogative.

Any given speaker's grammar could be a combination of any of these factors: thus, for example, there seemed to be no necessary link between the verb factors and the associate factors. Any factor appeared to be able to co-exist in a grammar with any other factor, indicating that they did not seem to derive from a single parameter setting or a cluster of parameter settings in the grammar.

4.   Implications for the nature of grammar

We have seen that a range of factors can have an impact on whether or not agreement occurs, and that impact is not the same for all speakers – the same factor which favours agreement in one speaker can disfavour it in another (for example, use of the past tense *was/were*). What we seem to be dealing with here is a case where, given variation in the input data, the acquired grammar can depend on a variety of factors. Moreover, these factors seem to be structure-specific: that is, speakers who allow lack of agreement with the associate in expletive constructions do not necessarily allow such lack of agreement in sentences without expletives.

There is, however, a certain naturalness to these processes. For example, the impact of position on agreement is well known from a language like Arabic (see for example Mohammad 1989), which has number agreement where the subject DP precedes the verb, but not where it follows it. The impact of animacy is well known, and in the particular context of plurality can be seen reflected in the Classical Greek 'neuter plural subject, singular verb' rule.[2] There are languages, such as Chinese, where the lexical item used for 'how many' differs when the expected answer is more or less than 10. What we seem to be seeing then is that, given variation in the input, the language faculty works out a grammar which fits most of the data, and that the factors which this grammar refers to can be many and varied (but probably constrained by the nature of the language faculty or language acquisition device). The process of Agree is a common one in language, though not quite universal (Chinese, for example, shows no agreement of any kind, at least overtly). What we seem to see here is that the language faculty is capable of working out rules and that such rules can refer to a range of properties of the elements of language. Our research goal should then be to look at the types of rules the language faculty is able to develop: we have found much about what is common to languages, now we need to find out the limits of how, at the individual level, they really differ.

There is, then, a multiplicity of possible grammars, some of them allowing lack of agreement in completely different circumstances to others, but all drawing on processes (such as agreement) and elements (such as sentence types and types of NP) which are well known to come into play in natural languages. Despite the wide variety of possible grammars, no speakers adopt rules referring to, for example, the numerical order of the verb in the sentence, or whether the associate contains a particular phoneme: processes which of course are not generally found in natural languages. These individual speakers therefore have adopted grammatical rules consistent with what is found in other language va-

rieties – these are not 'wild grammars'; but they have adopted a wide range of possible grammars.

On the contrary, Minimalist approaches to agreement generally see agreement as applying more or less blindly: a verb (or any other element) with an unvalued feature (such as person, number) acts as a *probe* and seeks a *goal* which shares a similar feature which is valued: thus a verb will seek for a noun which can value its agreement features (for recent accounts of agreement in the Minimalist framework, see Chomsky 2000, 2001b; Pesetsky & Torrego 2004). What is claimed to require this is that agreement features are seen to be uninterpretable, and uninterpretable features must be deleted because they cannot be semantically interpreted; they can only be deleted once valued. There is no reference here to features of the verb, or features of the element, such as the subject, triggering agreement. Rather, agreement is seen to be a universal property, acting on any element that emerges from the lexicon with unvalued features. If tensed forms of the verb *be* in general require to have their features valued, as they do in sentences where there is a full noun phrase in subject position, then they should also require to be valued in sentences with *there*, unless it is the case that *there* itself values the features of the verb. But the ability of *there* to value the features of the verb should not be dependent upon the features of the associate.

It is difficult to see how this, or any model grammar which seeks a very high degree of universality on classical minimalist lines can account in a descriptively, let alone explanatorily, adequate way for this type of grammatical variation. While some characteristics might be attributed to the properties of individual lexical items, and some to competing grammars, it is difficult to see how a pure minimalist account could account for grammars where the verb agrees with non-group-like associates obligatorily, provided they are in questions, but not necessarily in statements, and not necessarily where the verb is contracted on to the expletive, to give just one example of the grammars identified.

It could perhaps be argued that there are really only two grammars – with and without agreement – and that where these grammars are used is influenced by a range of essentially extra-grammatical factors; but if grammars are to characterize the internal linguistic knowledge of speakers, as exhibited in grammaticality judgements, then we must account for the grammars found, and for the fact that there are many of these.

Or it might be argued that agreement in sentences with expletive *there* is some kind of special case: Sobin (1997) claims that agreement in such sentences is, in fact, a 'grammatical virus', with non-agreement being the norm

and agreement with the associate being a learned, prescriptive rule. This argument is however opposed by Schütze (1998) who points out that agreement with the associate is entirely natural for many speakers of English, rather than having the claimed 'prescriptively taught' virus-like properties. In any case, the type of variation here is found in other, less 'virus-like' areas, for example in subject contact relatives. While Standard English does not have subject contact relatives (where the subject of the relative clause is relativized, and there is no use of an overt relative marker), many speakers allow subject contact relatives to a greater or lesser degree in colloquial English. A similar study to the one described in this paper, and currently in progress, indicates that the conditions under which subject contact relatives may occur appears to be just as variable across individual speakers as is the case with agreement in sentences with expletive *there*, depending variously on characteristics of the head noun, the verb in the main clause, and whether or not the verb in the subordinate clause is contracted. Thus some speakers allow subject contact relatives only if the head noun is an associate of an expletive. Others use them with any head noun which is a complement of the copula *be*, some others use them with *have* as well as *be,* and some a wider range of verbs including *see* and *meet;* for most speakers they are excluded where the head noun is in subject position, but others allow them with head nouns in all positions.

This paper then suggests that adequately characterising the nature and limits of human language grammars requires a return to a concept of rules and a research programme which seeks to identify and constrain the nature of those rules. This is not a step backwards, but rather, having gone a long way towards finding out what is universal in language, expanding on that to take account of the observed (probably constrained) degree of variation. Chambers (2003: 26–29) argues that it is a mark of maturity in a science to move beyond categoricity towards the study of variation, something he considers that sociolinguistics but not theoretical linguistics has done. We know that the important and seemingly universal characteristics of language are Merge, Move and Agree: we have seen that the circumstances under which at least the last occurs is more variable than might have been anticipated from a study which did not focus on syntactic variation among individuals.

Note that once we reintroduce the concept of rules, we reintroduce the possibility of those rules varying not only with linguistic characteristics but with extralinguistic ones; thus, for example, we can specify that certain rules only apply in formal usage, or that certain rules are 80% likely to apply under certain conditions. This enables the classic findings of sociolinguistics to be captured within the same architecture which is used in syntax. Sociolinguistic

variation, of course, is generally studied at the level of the community or group, rather than the individual; but at the level of the individual, it is likely that certain structures (agreement *versus* non-agreement, for example) also show frequency effects. Sociolinguists have been finding for a long time that the linguistic factors impacting on variables have been many and varied; they have found this by looking at recorded production at the community or group level. Paradoxically, it is by looking at the very micro-level, the syntax of individual speakers, that such variation shows up in using the methodology of intuitions.

Notes

1. *Associate* is the term generally used in generative syntax for the logical, postverbal subject in sentences with expletive *there.*

2. Of course it is not purely animacy that is at issue here: rather it is a conjunction of animacy with gender which determines agreement; in Classical Greek, inanimate nouns can be masculine, feminine or neuter. Only those grammatically neuter nouns which are also inanimate can have singular verbs; the small number of neuter nouns which are animate generally trigger plural agreement.

References

Chambers, Jack K. (2003). *Sociolinguistic Theory: Linguistic Variation and its Social Significance.* Oxford: Blackwell.

Chomsky, Noam (1957). *Syntactic Structures.* The Hague: Mouton.

Chomsky, Noam (1965). *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

Chomsky, Noam (1995). *The Minimalist Program.* Cambridge, MA: MIT Press.

Chomsky, Noam (2000). "Minimalist Inquiries". In Roger Martin, David Michaels, & Juan Uriagereka (Eds.), *Step By Step: Essays on Minimalist Syntax in Honor of Howard Lasnik* (pp. 89–156). Cambridge, MA: MIT Press.

Chomsky, Noam (2001a). *Beyond Explanatory Adequacy.* Cambridge, MA: MITWPL, 20.

Chomsky, Noam (2001b). "Derivation by Phase". In Michael Kenstowicz (Ed.), *Ken Hale: A Life in Language* (pp. 1–52). Cambridge, MA: MIT Press.

Haegeman, Liliane (1992). *Theory and Description in Generative Syntax: A Case Study in West Flemish.* Cambridge: Cambridge University Press.

Henry, Alison (1995). *Belfast English and Standard English: Dialect Variation and Parameter Setting.* New York: Oxford University Press.

Henry, Alison (in press). "Nonstandard Dialects and Linguistic Data". To appear in *Lingua.*

Kroch, Anthony (1989. "Reflexes of Grammar in Patterns of Language Change". *Language Variation and Change, 1*, 199–244.

Labov, William (1972). *Language in the Inner City: Studies in the Black English Vernacular.* Philadelphia: University of Pennsylvania Press.

Meechan, Marjory & Michele Foley (1994). "On Resolving Disagreement: Linguistic Theory and Variation – *There's Bridges*". *Language Variation and Change, 6,* 83–85.

Mohammad, Mohammad (1989). The Sentential Structure of Arabic. Ph.D. dissertation, University of Southern California.

Pesetsky, David & Esther Torrego (2004). "The Syntax of Valuation and the Interpretability of Features". Ms., MIT and U.Mass. Boston.

Schütze, Carson (1998). 'Expletive Constructions Are Not Infected". Ms., MIT.

Sobin, Nick (1997). "Agreement, Default Rules and Grammatical Viruses". *Linguistic Inquiry, 28,* 318–343.

Tortora, Christina (2003). *The Syntax of Italian Dialects.* New York: Oxford University Press.

Focus raising

A paradigmatic example of the treatment
of syntactic variation*

Judit Gervain and Gábor Zemplén

Scuola Internazionale Superiore di Studi Avanzati / Budapest University
of Technology and Economics, Hungarian Academy of Sciences

1.   Introduction: Defining the problem

Variation is a ubiquitous problem of the linguistic enterprise. It has often led
to one of two different approaches. One puts the collection and classification
of variation phenomena at the heart of the study of language, whereas the
other marginalizes it by offering no systematic treatment thereof. Sociolinguistics, in general, has taken the former path, while the latter is symptomatic of
biolinguistics.

Our paper is an attempt to investigate the status of variation in linguistics,
with special reference to syntactic microvariation. The questions we are asking
are mainly of a methodological nature: (1) what is the source of variation?, (2)
by what methods can it be detected and systematized?, and (3) how can the
variation patterns thus obtained be accounted for in a recognized theoretical
framework? By trying to establish an empirical method for the treatment of
variation, we subscribe to an emerging research agenda known as *experimental
syntax* (Schütze 1996; Cowart 1997; McDaniel & Cowart 1999; Keller 2000),
the aim of which is to place generative grammar on firm empirical grounds
through the use of better-controlled data gathering methods.

Throughout the paper, we will be assuming the Principles and Parameters
model of Chomsky (e.g. 1995, 2000). Under this view, the logically possible
space for human languages is defined by a set of universally valid principles
(applying to all natural languages) and a set of parameters or switches, the possible settings of which account for cross-linguistic variation. Even though we

will be using this particular model, our aim is to offer a novel tool for handling data that is compatible with a range of theoretical frameworks, as long as they are parametric in nature.[1]

First, we will describe the sources and types of variation and how they relate to current practices of data gathering procedures in biolinguistics. Second, we will go on to introduce and illustrate the new method. Third, we will conclude by discussing its advantages and limits.

2.   *Varietas delectat?* The problems, sources and types of variation

2.1   Delineating the problem

Let us first illustrate through concrete examples some of the problems that variation causes for current syntactic theory.

It is not uncommon to find native speakers of the same language who disagree about the well-formedness of a construction. Moreover, such controversies are present in the syntactic literature as well. Just to give a few examples, in an analysis of Hungarian possessive noun phrases, Alberti (2004) notes that the existing descriptions disagree about the grammaticality of the optional omission of the article. For Szabolcsi (1992), the article is obligatory in front of a possessive proper noun (see (1a) below, from Alberti 2004: example (17a)), while É. Kiss (1998) accepts the noun phrase without the article as well (see (1b) below, from Alberti 2004: example (17b)).

(1)   a.   *Eladtam   a/*Ø Péter   régi   kocsiját*
           Sell-1s-PST the   Péter-NOM old car-POSS3s-ACC
           "I sold Péter's old car."
      b.   *Eladtam a/Ø Péter régi kocsiját*

Another example comes from Hungarian focus-raising constructions, where É. Kiss (1987) and Lipták (1998) give conflicting generalizations about the case of the raised focus constituent. Focus-raising (FR) can be defined as a complex clause in which the focus constituent of the embedded clause surfaces in the matrix clause, as illustrated in (2).[2]

(2)   a.   *Azt   mondtad, (hogy) JÁNOST   láttad*
           expl-ACC say-2s-PST that   János-ACC see-2s-PST
           "You said that you saw János."
      b.   *JÁNOST$_i$ mondtad, hogy e$_i$ láttad*
           János-ACC say-2s-PST that   see-2s-PST

According to É. Kiss (1987:141), when a focused subject is raised, it can optionally bear Nominative (3b) or Accusative (3a) case. Lipták (1998:12–13), on the other hand, claims that Nominative case is ungrammatical (4a) in these constructions. It is interesting to note that Lipták bases her description on a small-scale survey in which she elicited judgements from a dozen native speakers, most of them linguists. Unfortunately, however, she does not give details as to the methods and procedures of data collection.

(3) a. *JÁNOS KÉT DOLGOT$_i$ szeretne,   ha t$_i$
       János two thing-sg-acc would.like-3s-indef if
       sikerülne
       succeed-cond
       "As for John, it is two things that he would like if they succeeded."

   b. *JÁNOS KÉT DOLOG$_i$ szeretné,   ha t$_i$
       János two thing-sg-nom would.like-3s-def if
       sikerülne
       succeed-cond

(4) a. **JÁNOS$_i$ mondtam, hogy t$_i$ megnyerte   a   versenyt*
       János-nom say-1s-pst that   win-3s-pst the race-acc
       "I said that John had won the race."

   b. *JÁNOST$_i$ mondtam, hogy t$_i$ megnyerte   a   versenyt*
       János-acc say-1s-pst that   win-3s-pst the race-acc

Although such inconsistencies often remain undetected, when they do arise, they give rise to considerable unease, as the following quote from Levine's review of Postal's *Three Investigations of Extraction* (2001:172–173) attests:

> But Postal's work raises questions of a somewhat darker nature. A possible response to at least the data-based challenge given in preceding sections is simply to accept that the null RP[=resumptive pronoun]-driven analyses exhibited in TIE[=*Three Investigations of Extraction*] are indeed untenable for speakers who accept these counterexamples as well-formed, but to maintain that the RP analyses correctly characterize the grammars of speakers who do not accept such data. [...] But logically, one story goes, that entails nothing about the grammar of other speakers. Even if Postal alone were to reject such counterexamples, would that not simply mean that the null RP hypothesis was defensible for at least one natural language grammar, *viz.* Postal's own?
>
> There is no simple answer to this question, which potentially arises in cases where investigators proposing different analyses of the same phenomenon disagree about the status of various crucial data.

Indeed, this question can lead to serious methodological dilemmas, which, are, however, rarely addressed. In this respect, Labov's (1975: 16) statement holds generally true, *viz.* "the logic of linguistic inquiry has been to assume consensus [about data] rather than test it".

But when the consensus cannot be upheld, the biolinguist arrives at an impasse. First, she rarely has enough information about the status of the incongruous data, since biolinguistic researchers usually remain reticent about the circumstances and methods of data gathering. Therefore, it is practically impossible to decide whether a given variation is geographical, social, idiolectal or is due to some other factor.

Second, when she intends to challenge a theory on an empirical basis, the unsolved data issues create an obstacle. Furthermore, rational theory choice is greatly hindered by conflicting data issues. For instance, since there is no way to choose between É. Kiss's and Lipták's data, it cannot be decided either which theory is empirically more adequate, thus preferrable. Also, how can one extend or improve a theory when one does not know what its empirical foundations are? What is the empirical scope and validity of a theory in the first place?

In a possible, but implausible *reductio ad absurdum*, as Labov (1975: 30, our emphasis) argues, "[i]f these intuitions are said to represent *only the linguist's idiolect*, then the value of his analyses rests on a very uncertain foundation." Although this scenario is not uncommon, it is undesirable, because it misses meaningful levels of generalization between the abstract Universal Grammar as a set of all the possibilities and the individual idiolects.

In sum, empirical issues may well create an obstacle for theoretical research, which is all the more serious, since some of these concerns go unnoticed or at least unmentioned, plaguing the very foundations of the enterprise.

2.2   Sources and types of variation

2.2.1   *The current practice of data collection in generative grammar*

How is it possible that a discipline like generative grammar, which praises itself on its descriptive and explanatory rigour and close resemblance to the hard sciences, allows such laxity with respect to its empirical bases? The answer lies in some of its core background assumptions.

A case in point would be Chomsky's introduction of the competence/performance (Chomsky 1972), and more recently the I(nternal)-language/E(xternal)-language distinction.[3] Competence or I-language is believed to be the language faculty in its abstract sense as instantiated in an individual's men-

tal grammar, i.e., the set of universal principles and the parameters in their specific settings. Performance or E-language is actual language use, influenced by language-external factors such as memory, world knowledge, social conventions etc. The main goal of generative grammar has been to describe linguistic competence or I-language. The most common method to access competence has been to study the grammaticality judgements of the idealized native speaker who is assumed to be able to abstract away from language external performance factors (e.g. Chomsky 1961). Thus, competence data are thought to be obtainable through native speaker introspection. Moreover, data sets are very often based on grammaticality judgments obtained either from the linguist herself or from a small number of informants, as opposed to groups or populations of speakers. This practice is based on the assumption that each individual has an instantiation of UG in her mind.

The idealization of the native speaker has led to a highly informal data gathering and analyzing practice.[4] In addition, statistical analyses are not carried out even when several informants are asked for judgements (Romaine 1980). However, other behavioural and brain sciences, especially psycholinguistics and cognitive science (Schütze 1996; Cowart 1997) have, in fact, shown that subjects are prone to non-linguistic biases in their linguistic perception and production, such as frequency effects, world knowledge factors etc. This may give rise to both noisy and potentially inconsistent data.[5]

Despite these issues, informal methodology is *not* intrinsic to the spirit of generative grammar, it does not logically follow from any of the fundamental commitments of the theory. As the following quote from Chomsky (1957: 13–14) illustrates:

> The grammar of L[anguage] will thus be a device that generates all of the grammatical sequences of L and none of the ungrammatical ones. One way to test the adequacy of a grammar proposed for L is to determine whether or not the sequences that it generates are actually grammatical, i.e., acceptable by native speakers, etc. We can take certain steps towards providing a behavioral criterion for grammaticalness so that this test of adequacy can be carried out.

Let us now consider what might be the prerequisites for taking such steps.

2.2.2 *Experimental tools and errors*
According to a not uncommon view in linguistics, the elicitation of native speaker judgements in generative linguistics, and the observation of linguistic production in sociolinguistics are similar in important ways to experimentation in natural sciences as a means of data gathering.

Most branches of the empirical sciences require tools for the systematic handling of 'observational error', and have accepted data-analysing and data-reducing procedures as a result. This is due to the fact that the data they operate with is not seen as obviously 'clean'. While this is not the general attitude in generative grammar, these considerations should naturally apply to this field as well, especially under the Chomskyan view (Chomsky 1972) that linguistics is a branch of the sciences of the human mind. Without such tools the criteria for accepting or rejecting some of the variants in the conflicting data in our introductory examples remain unclear and necessarily *ad hoc*. To use an analogy, this is like a chemist never bothering (or having no generally agreed-upon procedures) to check his equipment or the purity of his samples.

Thus it would seem fruitful for linguists to employ the concept of linguistic phenomena as separate from raw linguistic data. The model, originally developed for natural sciences (Bogen & Woodward 1988), attempts to bridge the gap between raw data and theory and disconnect the establishment of phenomena from the hypotheses it has a direct bearing upon. Applying this view to linguistics, theories should not predict data (i.e., an individual's grammaticality judgment of a sample sentence on a scale of 1 to 5), but linguistic phenomena. The individual data can be influenced by innumerable factors, and constructing a theory that predicts the particular items of data seems impossible (and unnecessary). Instead, the aim is to apply methods generally accepted by the scientific community to arrive at "stable, repeatable characteristics which will be detectable by means of a variety of different procedures, which may yield quite different kinds of data" (Bogen & Woodward 1988: 317). These phenomena are established on the basis of evidence from facts, they are systematized, unified, and result from a manageably small number of causal factors. While it remains impossible to obtain "theory-free" input for testing or developing specific hypotheses, this input is only influenced by notions commonly shared by scientists of the given field and not influenced by the individual and specific linguistic hypothesis in question. Thus the job of the linguist is threefold: (1) gathering data, (2) establishing phenomena on the basis of the data, and (3) finding hypotheses/theories that explain these phenomena. This implies that the analysis of the data and the establishment of phenomena have to be done independently of and prior to theoretical explanations (even if, of course, a theory of data processing has to be at hand). One of our main points is that this step has been mostly neglected in (generative) syntactic research.

These considerations bear directly on the problem of variation. As argued above, it is at the level of linguistic phenomena that empirical consensus must be sought, while actual facts are inherently variant, since they are influenced by

many different factors, some of which are external to language, and thus not directly relevant to linguistic theory. Linguistic methodology should therefore aim at recognizing, and, as far as possible, controlling these external factors, which, to continue the analogy with natural sciences, might be considered as sources of experimental error.

Recently, in a number of articles, Hon (e.g. 1989, 1998) has persuasively argued for a fourfold distinction of errors in the empirical sciences. His original typology – sketched below – has been developed on the basis of research practices in physics and other natural sciences, but once again, the analogy might shed light on some of the (as yet unrecognized) flaws of syntactic methodology noted above.

A first potential source of error may lie in the background theory, using a false general theory, false instrumental theories, or even a false theory of the experiment. Two issues are relevant here for linguistics. Firstly, as described earlier, generative grammar assumes a clear distinction between linguistic competence and performance. While accepting the distinction, we point out that the usual practice of informal data gathering relies on the false background assumption that competence is directly observable in performance. Psycholinguistics has provided ample evidence that it is not necessarily the case. Consequently, experimental controls are needed to filter out as much of the performance bias as possible. Secondly, difficulties also arise when online, real-time linguistic use and production are observed for their own sake, as in sociolinguistics. The problem called the observer's paradox is well-known in the variationist paradigm. Speakers tend to consciously manipulate certain aspects of their linguistic production when in the presence of a scientific observer or experimenter. This, of course, carries over to grammaticality judgements as well, since informants sometimes evaluate sentences on the basis of their 'school grammars' or with a prestigious dialect in mind.

Second, the assumptions concerning the actual set-up of the data gathering 'experiment' may be another source of error, e.g. faulty scaling, or the use of unstandarized instruments. In syntax, several considerations are in order in this respect. One important point concerns the grammaticality scale used for judgements. As Keller (2000) convincingly argues, grammaticality is very often graded, not binary, and distinguishing different levels of grammaticality is not only empirically more adequate, but also theoretically meaningful. A further point which is often raised (Schütze 1996; Cowart 1997) is the inclusion of linguistically trained subjects among the informants. A general argument against such a practice is that linguists are not naïve with respect to structural analysis, and can thus be unconsciously influenced by their theoretical commitments

when judging data. However, as Schütze (1996) and Cowart (1997) point out, empirical evidence has not confirmed the existence of such a bias. Our own results have shown no difference between naïve and linguist informants either (see Section 3.2 below).

Third, the actual observational phase can be a source of error, due to an incorrect marking or recording procedure or even to faulty reading of the data. This problem is most serious in linguistics when longer linguistic productions have to be transcribed (about transcription pratices, see Atkinson & Heritage 1999; Ochs 1999; Cameron 2001): hesitations, pauses, intonational patterns, false starts, corrections, deixis and communicative gestures are all very difficult to capture faithfully.

Fourth, the theoretical conclusions arising from the processing and interpreting of the recorded data can also lead to error. The result might be misunderstood or the interpretation might be erroneous.

Out of the four types of error, the first two are usually not recognized in linguistics. It is for this reason that we have argued for decoupling the evaluation of data from that of the linguistic theory (in so far as that is possible), and by doing so, establishing robust "phenomena" that can serve as the basis for theory testing and theory choice. We have also drawn attention to a possible typology of experimental error: the linguist's methods have to be devised in such a way as to reduce the risk of error.

2.2.3 *Competence variation*

It follows from the above that inconsistency and variation in the data may simply be the result of inadequate empirical methods. Experimentation, therefore, is a useful and necessary tool, but of course no definitive guarantee of less noisy, more consistent empirical data. However, performance and extralinguistic factors are not the only source of variation. Even if it could be ensured that a given data set is exempt from bias and 'noise', variation could still be present in the sample. The reason for this is that individual competences may vary. In the Principles and Parameters framework, this variation is not infinite or random, but falls out nicely and systematically from the possible combinations of the parameter settings. In other words, individual grammars vary just as different languages do. Crosslinguistic variation and interlinguistic microvariation are two of the possible and meaningful levels of generalization.

One case in point may be the variation between the Hungarian focus-raising constructions as described by É. Kiss (1987) and Lipták (1998), illustrated in (3) and (4) of Section 2.1. Since the authors are trained linguists, it is not unreasonable to suppose that they are less subject to performance factors

and other external biases, thus the difference in their generalizations is likely to reflect genuine competence differences. This hypothesis, however, has to be tested. This is what we will attempt in the remainder of the paper.

3.   An example: Detecting and analyzing microvariation in Hungarian focus-raising

So far, no attempt has been made in the generative paradigm to reliably detect and systematize competence variation.[6] Our objective in the following will be to introduce and illustrate a novel method for this purpose.

3.1   The data: Empirical problems

As shown above, the literature is not unequivocal with respect to an important property of focus-raising (FR) sentences.

The debated property, as noted earlier, is the case of the raised constituent. In É. Kiss's (1987: 141) description, a focus-raised subject may optionally exhibit either Nominative or Accusative case ((3) repeated as (5) here), whereas Lipták (1998: 12–13) argues, on the basis of judgements obtained from a dozen speakers, that only Accusative is admissible ((4) repeated as (6)).

(5) a. *János KÉT DOLGOT$_i$ szeretne,   ha t$_i$*
         János two   thing-SG-ACC would.like-3S-INDEF if
         sikerülne
         succeed-COND
         "As for John, it is two things that he would like if they succeeded."
     b. *János KÉT DOLOG$_i$ szeretné,   ha t$_i$ sikerülne*
         János two   thing-SG-NOM would.like-3S-DEF if   succeed-COND

(6) a. **JÁNOS$_i$ mondtam, hogy t$_i$ megnyerte   a   versenyt*
         János-NOM say-1S-PST that   win-3S-PST the race-ACC
         "I said that John had won the race."
     b. *JÁNOST$_i$ mondtam, hogy t$_i$ megnyerte   a   versenyt*
         János-ACC say-1S-PST that   win-3S-PST the race-ACC

In addition to this documented data controversy, another property of these constructions has been observed to vary across speakers. When the raised subject is a [$_{DP}$ [$_{NumP/QuantP}$ Num/Quant [$_{NP}$ Noun]]] type phrase, e.g. *két fiú* (lit.: two boy-SG, 'two boys'), for some speakers, the embedded verb may optionally agree with its subject in the singular (7a) or in the plural (7b), while other

speakers accept the canonical singular agreement only (since these DPs trigger singular agreement on their verbs in simple clauses (8)).[7, 8]

(7) a. *KÉT fiút$_i$ mondtál,   hogy t$_i$ jön*
         Two boy-SG-ACC say-2S-PST that   come-3s
         "You said that two boys were coming."
     b. *KÉT fiút$_i$ mondtál,   hogy t$_i$ jönnek*
         Two boy-SG-ACC say-2S-PST that   come-3P

(8) *Két fiú   jön/*jönnek*
     Two boy-SG-NOM come-3s/come-3P
     "Two boys are coming."

It is clear from the methodological discussions of the preceding sections that the theoretical analysis cannot proceed until these data issues are clarified. The problem is that it is impossible to decide whether variation is at the level of the facts or of the phenomena, thus whether it requires a (language) theoretical explanation or not. In fact, the two accounts give conflicting data to support two conflicting hypotheses. We have no grounds to reject either the data or the hypotheses – but also no arguments for accepting one or both of them. Three questions require answering in this regard:

a. Have measures been taken to recognize and possibly to rule out the different types of experimental error as described in 2.2.2 above?
b. If the answer to the question raised above is 'yes', despite our theory of error being unable to account for the differences and implying instead that they reflect differences in linguistic competence, what are the attested empirical patterns?
c. Is there any systematic relationship or covariation between the two controversial properties, i.e. case and agreement?

As for the first question, neither É. Kiss (1987), nor Lipták (1998) report on the methods by which their data was collected, but, as noted before, it is not implausible to assume that the two authors being linguists with trained grammatical intuitions, will have judgments that may reflect genuine competence differences. The experimental survey outlined below has been conducted by us with a view to addressing the other two issues.

3.2 An experimental solution

The first problem is to obtain clear competence data. The chance of experimental error both in the data gathering and the data processing phase must

be minimized. To achieve this goal, a paper-and-pencil questionnaire was used to elicit judgements from 23 native informants, some of them naïve, others linguistically trained.[9] The questionnaire contained both test sentences and fillers. The role of the latter was to counterbalance habituation effects, i.e., the indulgence in judgements that develops as informants get used to a given construction. Test constructions (e.g. FR of *két fiú* type subjects with plural embedded verbs, as in (9)) were always lexicalized in three different ways in order to avoid biases resulting from the different frequencies of lexical items, more or less plausible word combinations, pragmatic and world knowledge influences etc.

(9) a. *KÉT FIÚT   mondtál,   hogy jönnek*
         Two boy-sg-acc say-2s-pst that come-3p
         "You said that two boys were coming."
     b. *NÉGY SEBESÜLTET   jelentettek,   hogy meghaltak*
         Four   casualty-sg-acc announce-3p-pst that die-3p-pst
         "They announced (i.e. It was announced) that four casualties died."
     c. *ÖT SZÍNÉSZNŐT tudsz,   hogy öngyilkosok lettek*
         Five actress-sg-acc know-2s that suicide-pl become-3p-pst
         "You know that five actresses committed suicide."

Several construction types were created as a combination of a range of relevant linguistic factors, among which only case and agreement will be reported here for considerations of space. The possible combinations of all the factors and the different lexicalizations thereof yielded 53 test sentences, which were randomly intermixed with the 12 filler sentences. The full questionnaire was given to subjects with detailed instructions as to the intended pronunciation and interpretation of focus constructions together with clear examples illustrating grammaticality judgements (vs. falseness or semantic implausibility). In addition, informants were instructed to rely on their intuitions and not their "scholastic grammatical knowledge". It can reasonably be assumed that speakers had no *a priori* attitude with respect to FR sentences or with respect to what is expected as the "better speech" answer, because FR constructions are unmentioned in prescriptive or school grammars,[10] and, except for maybe a handful of linguists, average speakers, although they use it frequently, are not conscious of the existence of this structure.

Subjects were asked to evaluate the grammaticality of each sentence on a five-grade scale ranging from totally ungrammatical (*) to fully grammatical (OK) through three intermediate levels (???, ??, ?). A five-grade scale was adopted to allow comparison with other studies, since, if graded grammatical-

ity scales are used at all in syntactic analyses, five is the most common number of levels. More details about the experimental procedure and the full material are available in Gervain (2002), however, the instructions and their English translation appear in the Appendix to the current paper.

As a second step, informants' responses were quantified and given statistical treatment. This is the move from the experimental data to phenomena, or step (2) as described in 2.2.2 above. Note that no specific linguistic hypothesis about the nature of focus-raising in Hungarian affects this level. Therefore this level is (i) independent of the actual hypothesis testing, (ii) can also be criticized independently and (iii) consensus can be reached, even if the explanatory hypotheses differ markedly.

First, it was established that the variance is indeed statistically significant for some of the test sentences.

Second, a classificatory method called cluster analysis (Lance & Williams 2001; Everitt 1981, 1996; Van Ooyen 2001) was carried out to systematize the responses. Cluster analysis is a collection of 'heuristic' methods for the categorization of objects according to some similarity measure along one or, typically, several characteristics (variables). It is widely used, for instance, in microbiology to establish different strains of bacteria, or in biological taxonomy, to set up species, families, genera etc. of organisms. Cluster analysis has different types and techniques depending on the similarity measure and the classificatory criterion being used, but the underlying logic is the same – the stepwise, reiterative grouping of the two most similar objects or already formed clusters. Thus, agglomerative clustering starts out with as many clusters as there are objects in the sample, and, by repeatedly putting together those two items (two objects, an object and a cluster, or two clusters) that are the most similar with respect to the measured characteristics, it gradually decreases the number of clusters until all objects eventually belong to one big category. (Divisive clustering proceeds in the other direction.) The result of the categorization is plotted on a dendrogram (e.g. Figure 1), from which the classes can be read off.

Similarity is measured as a kind of 'distance' in the character space, an n-dimensional space defined by the n number of variables (properties, characteristics) observed in the experiment. Without attempting to give an exhaustive list of these measures, we will simply introduce the most frequently used ones. First, the measure called Euclidean distance is determined as the actual physical distance between two points in the n-dimensional space. A second possibility for measuring similarity, known as the city block or Manhattan distance, is to add up the distances along all the axes in the n-dimensional space. Besides these

two well-established measures, many others are reported in the literature (e.g. Van Ooyen 2001).

Once the distance between the original objects is determined, the two closest, i.e. most similar, can be fused to form a cluster. However, clusters are not 'natural objects' in the data set. They have no values of their own for the observed variables. These have to be obtained from the original objects that are included in each cluster. This calculation is carried out following a clustering algorithm. Again, several algorithms exist in the literature, and only the most popular ones will be introduced here. The Unweighted Pair-Groups Method Average (UPGMA) defines the distance between a cluster, on the one hand, and an object or another cluster, on the other, as the average of the distances between each point in the cluster and the object or each point of the other cluster. The Furthest Neighbor method determines the distance between two clusters or a cluster and an object as the greatest distance between a member of the first and the second cluster, or the greatest distance between the object and a member of the cluster. Since distances are maximized, this method tends to produce well-defined, sharply distinguished clusters. The Nearest Neighbor method, on the other hand, chooses the smallest distance, and produces rather elongated clusters standing closely together, a reason for which the method is also known as chaining. A fourth technique, Ward's method fuses those items into a cluster, the fusion of which augments within-cluster variance the least. This procedure also results in strongly demarcated groups, but also lumps together the outliers that have nothing in common except that they are different from everything else.

As pointed out before, clustering is a heuristic method. It thus requires some more or less subjective decisions on the part of the user. The first one is whether to use clustering at all. The crucial assumption here, which justifies the use of clustering, is that linguistic microvariation is not random, but is assumed to follow the patterns that result from the different settings of the parameters. The validity of this assumption is granted by the general parametric framework (e.g the Principles and Parameters model) we assume. The second decision that has to be made when clustering is applied concerns the similarity measure and the clustering algorithm that are chosen. Here we used Euclidean distance because it is the most simple and straightforward measure of similarity, and Ward's clustering algorithm.

The results of clustering appear in Figure 1. The closer two informants are linked up in the tree, the more similar their responses along the linguistics factors tested in the survey.
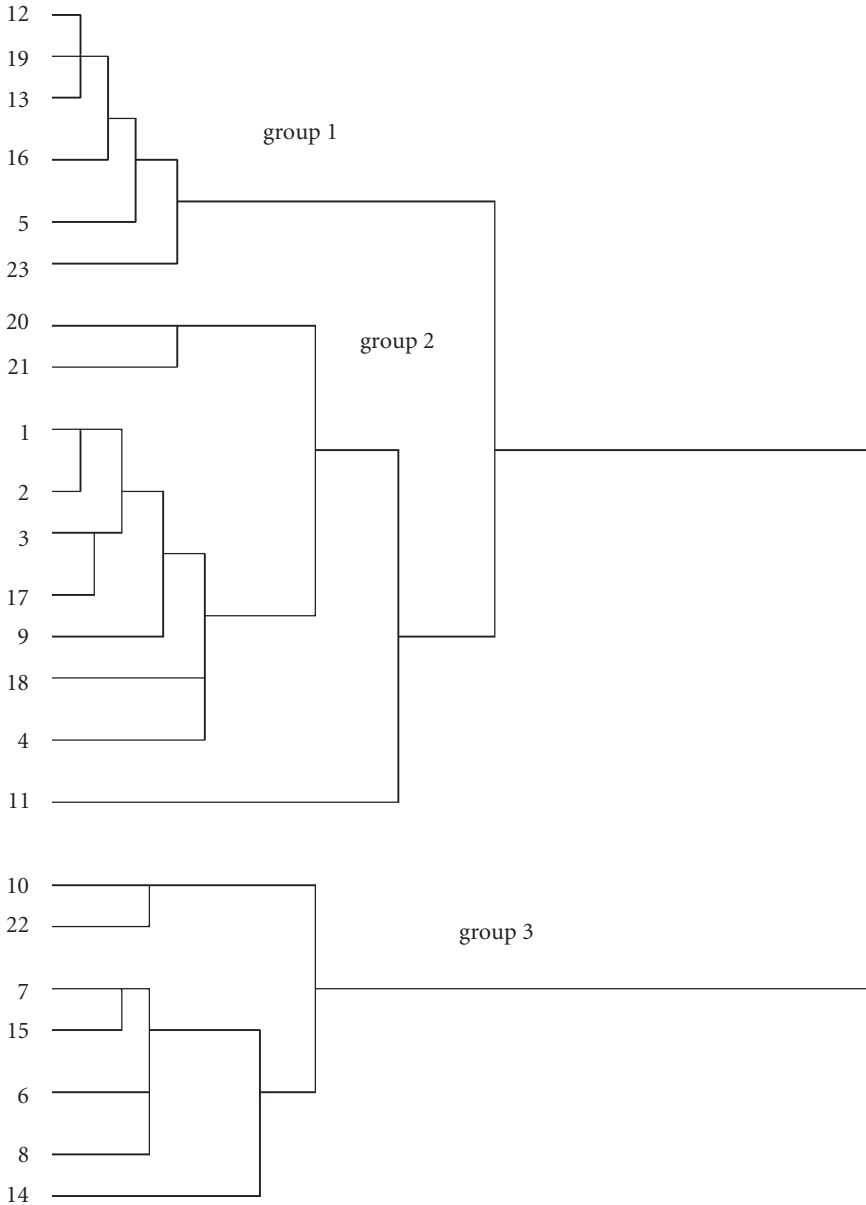
Figure 1. The dendrogram obtained by cluster analysis. Numbers indicate individual speakers.
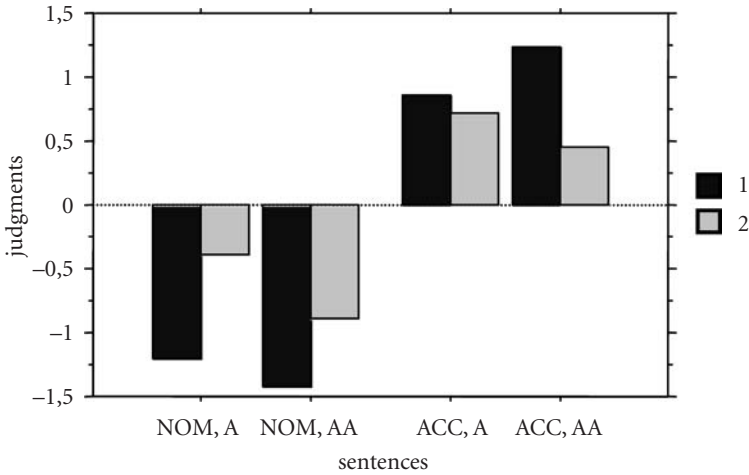
Figure 2. The results of the analysis of variance for the different sentence types. NOM: Nominative case, ACC: Accusative case, A: proper agreement (singular), AA: anti-agreement (plural).

Three clusters or groups were identified, as indicated on Figure 1. One major divide distinguishes groups 1 and 2 on the one hand, and group 3 on the other. While informants belonging to the first two groups accept some form of FR, group 3 speakers reject it altogether. A second main distinction differentiates between group 1 and group 2 (with two outliers, individuals nr. 4 and 11).[11] A two-way analysis of variance (ANOVA) was carried out to test the differential effects of the two decisive factors, case and agreement, in the two groups.[12] The following statistical results were obtained (Figure 2).

A highly significant main effect of case ($F(1,12) = 246.788$, p < .0001) was found. Agreement and group membership showed no main effect ($F(1,12) = 2.667$, ns; $F(1,12) = .251$, ns, respectively). Significant interactions were obtained for case × group membership ($F(1,12) = 24.473$, p < .001), for agreement × group membership ($F(1,12) = 6.461$, p < .05) and for case × agreement ($F(1,12) = 6.067$, p < .05). No triple interaction was attested ($F(1,12) = 1.105$, ns).

These results reflect the following linguistic protocols. Group 1 speakers refuse Nominative case with either agreement pattern, but accept Accusative with both (10), whereas group 2 informants accept singular, i.e. regular agreement, with both cases, but judge plural, i.e. irregular agreement as ungrammatical whatever the case of the focused constituent (11). It has to be noted here that these results obtained and protocols hold true of speakers not only as

a group, but also of each individual separately. In other words, each and every speaker in group 1 exhibits the pattern in (10), while all informants in group 2 show the profile given in (11).

(10) a. ???*AZ ÖSSZES LÁNY*   *mondtad*, *hogy jön*
          The all   girl-SG-NOM say-2S-PST that   come-3S
          "You said that all the girls were coming."

      b. *AZ ÖSSZES LÁNY   *mondtad*, *hogy jönnek*
          The all   girl-SG-NOM say-2S-PST that   come-3P

      c. ?*AZ ÖSSZES LÁNYT*   *mondtad*, *hogy jön*
          The all   girl-SG-ACC say-2S-PST that   come-3S

      d. *AZ ÖSSZES LÁNYT*   *mondtad*, *hogy jönnek*
          The all   girl-SG-ACC say-2S   that   come-3P

(11) a. ?*AZ ÖSSZES LÁNY*   *mondtad*, *hogy jön*
          The all   girl-SG-NOM say-2S-PST that   come-3S
          "You said that all the girls were coming."

      b. *AZ ÖSSZES LÁNY   *mondtad*, *hogy jönnek*
          The all   girl-SG-NOM say-2S-PST that   come-3P

      c. *AZ ÖSSZES LÁNYT*   *mondtad*, *hogy jön*
          The all   girl-SG-ACC say-2S-PST that   come-3S

      d. **AZ ÖSSZES LÁNYT*   *mondtad*, *hogy jönnek*
          The all   girl-SG-ACC say-2S-PST that   come-3P

Thus, both É. Kiss's (1987) and Lipták's (1998) position can be made sense of in the following variationist description. É. Kiss's (1987) 'dialect' corresponds to the group 2 protocol, while Lipták's (1998) generalizations to group 1. Furthermore, it has been shown that there is a systematic relationship between the two controversial factors, i.e. case and agreement, since only two of the four possible combinations are attested. Variation has been shown to exist at the phenomenal level, i.e. it is robust, reliable, replicable and exists accross speakers.

   The question to be asked at this point is how to account for the two distinct patterns: how can we make theoretical sense of the variation?

3.3 Towards a theoretical analysis

No detailed analysis will be offered here (for a more complete theoretical discussion, see Gervain 2002). Rather, the aim of this section will be to illustrate how generative syntactic theory, usually believed to be unable to handle optionality, and thus variation, can account for the obtained empirical pattern.

Previous analyses (É. Kiss 1987; Kenesei 1994; Lipták 1998), though different in their details, all analyze FR as Movement. The focused constituent is moved from its embedded position to the matrix focus phrase. While an account of this kind is able to explain the variant in which the verb agrees with the focused subject in the singular (group 2), plural agreement cannot be derived, since, as (8) shows, a *két fiú* 'two boy' type DP never triggers plural agreement on the verb in simple clauses. Therefore, the DP cannot start out as the subject of the embedded clause when the embedded verb is in the plural. Another strategy is needed to account for the protocol attested in group 1. The focused DP must be base-generated directly in the matrix clause, while the embedded subject position is filled with a dummy, phonologically null resumptive pronoun, which enters into a dependency relationship with the focused DP. Since these DPs are morphosyntactically, i.e. in intrasentential agreement, singular, but semantically, i.e. in intersentential anaphora, plural (they are referred back to intersententially with a plural pronoun, as in (12) – note that this overt pronoun is *not* a resumptive, but a fully referential ordinary pronoun), the resumptive pronoun may inherit either of the number properties through the resumptive dependency.

(12) *Két fiú$_i$ érkezett.   Nem ismertem fel *őt/őket$_i$*
      Two boy-SG-ACC arrive-3S-PST NEG recognize PRT him/them
      "Two boys arrived. I did not recognized them."

Thus the resumptive strategy is able to derive both the singular and the plural agreement.

Case facts also follow. In the movement strategy, the focused DP is assigned Nominative case in the embedded subject position, and as it moves, it also receives Accusative case from the main verb. In this case conflict (Español-Echevarría & Ralli 2000), either of the two structural cases may surface, hence the optionality of case in the movement (group 2) data. In the resumptive derivation, the DP is only assigned one case, the Accusative of the matrix verb, therefore it is obligatory.

One issue remains, however. If Chomsky (1981, 1982) is right in assuming that resumption is a 'last resort' strategy, that is, it can only be made use of when other derivations, especially movement are barred, why is it the case that some speakers employ the resumptive strategy rather than movement? Why is movement blocked for these subjects? Hungarian offers no empirical evidence to answer this question. Shlonsky's (1992) analysis of resumptives in Palestinian and Hebrew is, however, suggestive in this respect. Palestinian uses different complementizers for embedded clauses that contain resumptives than

for those that contain traces of movement. In a similar vein, we assume that although the Hungarian complementizer is phonologically identical in both cases, there are actually two lexical items corresponding to the complementizer, one of which blocks movement. The lexical choice of the complementizer for speakers who construe FR through resumption is different from that of the other group of speakers. Further research is of course needed at this point, since (i) the relevant empirical evidence has to be found in Hungarian, and (ii) from a theoretical point of view, it has to be explained how this relates to Rizzi's Relativized Minimality constraint, according to which the complementizer should not block A'-movement, since it is a head. The fine structure of the left periphery of the embedded CP in Hungarian will have to be examined in more detail. However, on the basis of the Hebrew and Palestinian data, it is not unreasonable to assume that variation is ultimately a question of lexical choice here.

4.   Discussion and conclusion

The generative approach, usually characterized as a hypothetico-deductivist paradigm (Allan 2003), has paid very little attention to developing sophisticated data-gathering and evaluating techniques. Chomskyan linguistics is little equipped to interpret "unclear" data it encounters.

Our article has shown that it is possible to do syntax with a watchful eye on potential sources of experimental error, thus augmenting empirical adequacy. The final results of the syntactic analysis relate, therefore, to the methodological considerations in two ways. We have shown that (i) the difference between É. Kiss and Lipták is *not*, as far as it can be experimentally established, random or extraneous to language, so (ii) it *is* a true dialect split, for the empirical treatment of which we offer a novel tool, cluster analysis.

This new approach increases consciousness about linguistic phenomena as different from both facts and theory. The ability to establish robust linguistic phenomena allows us to solve the ever increasing number of cases where consensus about the data was previously unreachable, such as the Hungarian case provided above.

The new level of description that emerges, i.e. that of linguistic phenomena, can serve as a common and negotiable empirical ground for several different approaches to the study of syntax. In sociolinguistics, new ideas have been put forth for the conceptualization of syntactic variation, partly adopting the Principles and Parameters model of generative grammar. Under this view (e.g.

Cornips 1998), syntactic variants only relate to social factors indirectly; thus, at an I(nternal)-linguistic level, they are considered to be differences in parameter setting, which, on a social level, are attributable to E(xternal)-variation.

At the intersection of these new tendencies, there seems to be emerging a new framework (Cornips & Corrigan 2005) that is able to capture both the internal and the external aspects of syntactic variation. This framework remains paramteric in the Chomskyan (Chomsky 1995) sense, instantiating a discrete rule system conceived of as a mental faculty of the individual speakers. At the same time, it opens up the way to the study of how these individual rule systems are ultimately related to social factors.

Notes

1. An emerging new approach (Manning 2003) has recently made the claim that syntax is genuinely non-categorical and non-parametric, assuming that probabilistic models can best describe the facts of variable language use. Such a theory is undesirable in our view, because (i) from a methodological point of view, it very often aggregates over speakers (using large corpora instead of individual informants), and (ii) from a theoretical point of view, because it does not offer a principled explanation for the lack of certain structures (it assigns zero probability to them, but does not say why it is that they cannot exist). Therefore, it is unsuitable for detecting small, but systematically and parametrically-occurring variations, which may, however, be relevant for our understanding of UG, as will be shown below.

2. Small capitals in the Hungarian examples below indicate lexical items that bear focal stress. The English translations do not render the interpretation of focus by grammatical means (e.g. clefts). Rather, the English equivalent of the Hungarian focus constituent should be pronounced with adequately strong, i.e. focal or main stress.

3. The two dichotomies will be used interchangeably here, though certain authors do make a distinction between them.

4. For a criticism of introspective methodology, see Milroy (1984, 2001).

5. The most common of these biases are pragmatic and lexical effects, including frequency (e.g. more frequent lexical items and combination may be judged as more tolerantly), semantic and world knowledge effects (e.g. true sentences and those that denote known states of affairs may be evaluated more indulgently), memory limitations (e.g. highly complex con-

structions may be rejected) and habituation (e.g. when a structure is encountered repeatedly it tends to be judged more and more grammatical).

6. Some accounts (e.g. Den Dikken 1999) do distinguish and analyze 'syntactic microdialects'. However they do not report on the methods used to establish the variants and their treatment of the data remains, in general, very informal.

7. Note that the noun can never be marked for plural, i.e., *két fiúk* (two boy-PL) is completely ungrammatical.

8. For the sake of simplicity, examples, at this juncture, are given only with Accusative case on the raised subject. The interaction of case and agreement patterns will receive ample treatment later on, though.

9. The statistical analyses revealed no significant difference between the responses of the two groups, $F(1, 21)= .846$, ns.

10. Actually, until the first description by É. Kiss (1987), we know of only one traditional, descriptive (but not prescriptive) grammar (Zolnay 1926) that makes any mention of FR at all.

11. These informants were rejected because their protocols were inconsistent, deviating from the rest of the speakers even with respect to the filler sentences that served as a control, since no variation was otherwise attested for these sentences.

12. The analysis of variance was carried out to further strengthen the results found by the cluster analysis. Importantly, it is not employed to validate them. Clustering is a statistical method in its own right, and even if its results fail to reach significance in a traditional ANOVA, they have a value of their own.

Nevertheless, an additional survey was carried out later for independent reasons. In this second study (Gervain 2004), the informants who have previously been established as belonging to group 1 and 2 respectively were given new FR constructions to judge (in the familiar paper-and-pencil paradigm). The ANOVA applied to this new and independent data set showed a statistically highly significant difference (*t*-test, $p < .0001$) between the two groups, thus confirming the results reported in this paper.

References

Alberti, Gábor (2004). "Az Eltűnt Névelő Nyomában". In Márta Maleczki & László Büky (Eds.), *A Magyar Nyelv Leírásának Újabb Módszerei VI*. Szeged: JATEPress.

Allan, Keith (2003). "Linguistic Metatheory". *Language Sciences, 25*, 533–560.

Atkinson, J. Maxwell & John Heritage (1999). "Jefferson's Transcription Notation". In Adam Jaworski & Nikolas Coupland (Eds.), *The Discourse Reader* (pp. 158–166). London: Routledge.

Bogen, James & James Woodward (1988). "Saving the Phenomena". *The Philosophical Review, XCVII*(3), 303–352.

Cameron, Deborah (2001). *Working with Spoken Discourse*. London: Sage.

Chomsky, Noam (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam (1961). "Some Methodological Remarks on Generative Grammar". *Word, 17*, 219–239.

Chomsky, Noam (1972). *Language and Mind*. New York: Harcourt.

Chomsky, Noam (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam (1982). *Some Concepts and Consequences of the Theory of Government and Binding* (Linguistic Inquiry Monographs 6). Cambridge, MA: MIT Press.

Chomsky, Noam (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Cornips, Leonie (1998). "Syntactic Variation, Parameters, and Social Distribution". *Language Variation and Change, 10*, 1–21.

Cornips, Leonie & Karen Corrigan (2005). "Convergence and Divergence in Grammar". In Peter Auer et al. (Eds.), *Dialect Change: Convergence and Divergence in European Languages* (pp. 96–134). Cambridge: Cambridge University Press.

Cowart, Wayne (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks: Sage Publications.

Den Dikken, Marcel (1999). "On the Structural Representation of Possession and Agreement". In István Kenesei (Ed.), *Crossing Boundaries: Advances in the Theory of Central and Eastern European Languages*. Amsterdam: John Benjamins.

É. Kiss, Katalin (1987). *Configurationality in Hungarian*. Budapest: Akadémiai Kiadó.

É. Kiss, Katalin (1998). "Mondattan". In Ferenc Kiefer, Katalin É. Kiss, & Péter Siptár (Eds.), *Új Magyar Nyelvtan* (pp. 156–168). Budapest: Osiris.

Español-Echevarría, Maria & Angelica Ralli (2000). "Case Mismatches in Greek: Evidence for the Autonomy of Morphology". *Acta Linguistica Hungarica, 47*, 179–203.

Everitt, B. (1981). *Cluster Analysis*. London–New York: Heinemann–Halsted.

Everitt, B. (1996). *Making Sense of Statistics in Psychology*. Oxford: Oxford University Press.

Gervain, Judit (2002). Linguistic Methodology and Microvariation in Language. MA Thesis. Szeged: University of Szeged.

Gervain, Judit (2004). "Hungarian Focus-Raising Revisited: Syntactic and Semantic Considerations". In Benjamin Shaer, Werner Frey, & Claudia Maienborn (Eds.), *ZAS Papers in Linguistics: Proceedings of the 'Dislocated Elements in Discourse' Workshop, Berlin, 18–20th December 2003* (pp. 111–132). Berlin: ZAS.

Hon, Giora (1989). "Towards a Typology of Experimental Error: An Epistemological View". *Studies in History and Philosophy of Science, 20*, 469–504.

Hon, Giora (1998). "'If This Be Error': Probing Experiment With Error". In Michael Heidelberger & Friedrich Steinle (Eds.), *Experimental Essays – Versuche zum Experiment* (pp. 227–248). Baden-Baden: Nomos.

Keller, Frank (2000). Gradience in Grammar. Ph.D. dissertation. Edinburgh: University of Edinburgh.

Kenesei, István (1994). "Subordinate Clauses". In Ferenc Kiefer & Katalin É. Kiss (Eds.), *The Syntactic Structure of Hungarian. Syntax and Semantics* (pp. 275–354). New York: Academic Press.

Labov, William (1975). *What is a Linguistic Fact?* Lisse: The Peter de Ridder Press.

Lance, G. N. & W. T. Williams (2001). "A General Theory of Classificatory Sorting Strategies. 1. Hierarchical systems". *Computer Journal, 9*, 373–380.

Levine, Robert (2001). "The Extraction Riddle: Just What Are We Missing?" *Journal of Linguistics, 37*, 145–174.

Lipták, Anikó (1998). "A Magyar Fókuszemelések Egy Minimalista Elemzése". In László Büky & Márta Maleczki (Eds.), *Proceedings of A Mai Magyar Nyelv Leírásának Újabb Módszerei III* (pp. 93–115). Szeged: JATE Press.

Manning, Christopher D. (2003). "Probabilistic Syntax". In R. Jens Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic Linguistics*. Cambridge, MA: MIT Press.

McDaniel, Dana & Wayne Cowart (1999). "Experimental Evidence for a Minimalist Account of English Resumptive Pronouns". *Cognition, 70*, B15–B24.

Milroy, James (2001). "Language Ideologies and the Consequences of Standardization". *Journal of Sociolinguistics, 5*(4), 530–555.

Milroy, Lesley (1984). "What a Performance! Some Problems with the Competence-Performance Distinction". In R. Hawkins & J. Harris (Eds.), *Sheffield Working Papers in Language and Linguistics* (pp. 40–52). Sheffield: University of Sheffield.

Ochs, Elinor (1999). "Transcription as Theory". In Adam Jaworski & Nikolas Coupland (Eds.), *The Discourse Reader* (pp. 167–182). London: Routledge.

Romaine, Suzanne (1980). "The Methodology of Urban British Sociolinguistics". *English World-Wide, 1*, 163–198.

Schütze, Carson (1996). *The Empirical Base of Linguistics. Grammaticality Judgments and Linguistic Methodology*. Chicago: Chicago University Press.

Shlonsky, Ur. (1992). "Resumptive Pronouns as a Last Resort". *Linguistic Inquiry, 23*(3), 443–468.

Szabolcsi, Anna (1992). *A Birtokos Szerkezet és az Egzisztenciális Mondat.* Budapest: Akadémiai Kiadó.

Van Ooyen, Arjen (2001). "Theoretical Aspects of Pattern Analysis". In L. Dijkshoorn, K. Towner, & M. Struelens (Eds.), *New Approaches for the Generation and Analysis of Microbial Fingerprints* (pp. 31–45). Amsterdam: Elsevier.

Zolnay, Gyula (1926). "Mondatátszövődés" [Sentence Intertwining]. *Értekezések a Magyar Tudományos Akadémia Nyelv- és Széptudományi Osztálya Köréből*, 23.

Appendix: Instructions to the questionnaire

The original Hungarian version:

Jók-e nyelvtani szempontból a következő mondatok? Mindig a megérzésére hallgasson, ne az iskolában tanultakra!

A nyelvtanilag jó mondatok nem kell, hogy jelentésüket tekintve igazak, helyesek vagy valóságokat legyenek.

A nagybetűvel írt szavak hangsúlyosak a mondatban. Pl.:

(1)   ANNA érkezett meg. (értsd: nem valaki más)

Ha a mondat jó, semmilyen jelet ne tegyen a mondat előtti kis négyzetbe.

(2)   ANNA írta meg a levelet.

Az egyértelműen rossz mondatokat csillagozza meg. Pl.:

(3) *ANNA írt meg a levelet.

Ha bizonytalan ítéletében, akkor a mondat jóságának megfelelően tegyen egy, két vagy három kérdőjelet a kis négyzetbe. Pl.:

(4) ??Anna akarta megírni kezdeni a levelet.

Köszönöm a segítséget!

The English translation:

Are the following sentences well-formed as far as the structure is concerned? Always base your judgment on your intuition, not on what you have learnt at school.

Sentences with well-formed structures need not be true, and their meanings need not be adequate or plausible.

Words that appear in capital letters are to be pronounced with emphasis.

(1) ANNA has written the letter. (meaning not someone else)

If the sentence is well-formed, do not put any sign in the box preceding the sentence.

(2) ANNA has written the letter.

If a sentence is undoubtedly ungrammatical, mark it with an *.

(3) *ANNA written a letter.

If you are uncertain or feel that the sentence is somewhere in between the two extremes, put one, two or three question marks (?, ?? or ???) in the box.

(4) ?ANNA wanted to be able to start to write the letter.

Thank you very much for your help.

Syntactic variability, social stratification and real/apparent time

Variation and the minimalist program*

David Adger and Jennifer Smith
Queen Mary, University of London / University of York

1.   Introduction

Granting the presence of widespread structured variation in the use of language (e.g. Labov 1972a), the following question arises for theoretical linguistics: how is the mental grammar (the I-language of Chomsky 1986) organized so that such variation arises?

There have been a number of kinds of answers to this question. An early response within the variationist paradigm was to posit variable rules (e.g. Labov 1972b; Cedergren & Sankoff 1974), where probabilities are built into the definition of the grammatical rules themselves. Other approaches involve positing multiple grammars, or the idea that multiple parametric settings are available to speakers (see Bender 2001; Henry 1995; Kroch 1989a, 1989b for different approaches).

In this paper, we argue that the architecture of the Minimalist Program (Chomsky 1995, 2000) is particularly well suited to dealing with grammatical variability, because of the way that it manipulates grammatical features, essentially allowing variable phonological outputs with the same semantic interpretation. We illustrate this approach via two case studies from a Scottish dialect (Smith 2000).

2.   The framework

The framework of the Minimalist Program (Chomsky 1995, 2000, 2001) proposes that knowledge of language can be captured as a function from sets of lexical items to meaning-sound pairs. This function is defined by a small number

of very general syntactic operations. The syntactic operations themselves are uniform across the species, with all cross-linguistic variability being confined to the specification of lexical items. Lexical items themselves are considered to be just collections of features: morpho-phonological, semantic and syntactic. The framework assumes that syntactic features come in two sorts: those with a semantic interpretation (such as the feature [past]), and those with a purely syntactic function (for example, nominative case on a nominal, or agreement marking on a verb). The former features are called *interpretable*, while the latter are *uninterpretable*. The way that uninterpretable features capture syntactic dependencies is that they have the following property: an uninterpretable feature must be *checked* by a matching feature (whether interpretable or not). This means that every syntactic dependency will be triggered by the presence of an uninterpretable feature.

We will follow the implementation of Minimalism developed in Adger (2003), for concreteness. In this implementation, features are assumed to generally have the form [Feature:Value] (e.g. [case: nominative]). Uninterpretable features on lexical items may lack a value (e.g. we notate uninterpretable features by prefixing them with a *u,* following Pesetsky & Torrego 2001, so, for example, the lexical entry for a noun might just have the specification [*u*case:]). Such lexical items receive their value during the syntactic derivation. This happens when they check with a matching feature which does have a value. Schematically: a structure containing [F:a] ... [*u*F:] becomes one which contains [F:a] ... [*u*F:a].

Once an uninterpretable feature is checked, it deletes. We notate this with a strikethrough, again following Pesetsky and Torrego's notation, so, refining what we saw immediately above, we have a structure containing [F:a] ... [*u*F:] becoming one containing [F:a] ... [~~*u*F~~:a].

The result of this system is that the final representation delivered to the semantic component consists only of interpretable features; all uninterpretable features have been deleted.[1]

To see how this works in a concrete case, take a sentence like (1), where the subject bears two features: [number:plural], an interpretable feature which we abbreviate as [num:pl], and [*u*case:nominative], an uninterpretable one, which we abbreviate as [*u*case:nom].

(1)   They bark

We assume that the whole sentence is headed by a syntactic element T, which in this case bears an interpretable present tense feature [tense:pres], and two different uninterpretable features: [*u*case:nom] and [*u*num:pl]. This specification

captures the fact that nominative case in English is syntactically dependent on finite T and also the fact that English verbs agree in number (we ignore first person singular agreement with the present form of *be* in English, here. See Section 5 for discussion). Following much work on English verbal morphology (see, for example, Chomsky 1957; Lasnik 1981; Bobaljik 1995) we assume that the morphology on the main verb arises because a morphological operation has spelled out the tense and number features of T as a verbal suffix.[2]

The system relates the sound pattern of the orthographical representation in (1) to its meaning by means of a syntactic derivation like that in (2).

(2) a. Select relevant lexical items:{*pronoun*[num:pl, *u*case:], *bark*[V], T(ense)[tense:pres, *u*case:nom, *u*num:]}.[3] Note that the case feature on the noun is unvalued, as is the number feature on T.

   b. Group together *pronoun* and *bark*, creating a VP constituent. The grouping operation is called *Merge*.

   c. Merge Tense with VP creating a T(ense) P(hrase):

<pre>
                        TP
                       / \
     T[tense:pres, ucase:nom, unum:]   VP
                     | |
        pronoun[num:pl, ucase:]   bark
</pre>
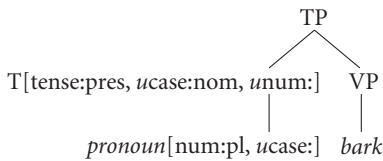
   d. Set up a checking relation between T and *pronoun* in the new structure, which checks and values the uninterpretable case and number features on both. This checking relation is usually called *Agree*. Since both case features are uninterpretable, both are marked with a strikethrough. Since, of the two plural features in the representation, only [*u*num:] on T is uninterpretable, only it is marked with a strikethrough:

<pre>
                        TP
                       / \
     t[tense:pres, ~~ucase~~:nom, ~~unum~~:pl]   VP
                     | |
        pronoun[num:pl, ~~ucase~~:nom]   bark
</pre>
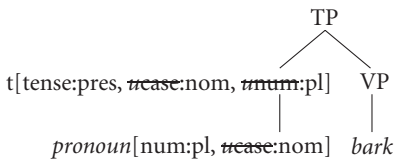
   e. Move the pronoun into the specifier of Tense Phrase, leaving a copy in the original position, notated here with angled brackets (we ignore the motivation for this movement here, see Chomsky 1995). This operation is called *Move*.

TP
pronoun[num:pl, ~~ucase~~:nom]   T'
T[tense:pres, ~~ucase~~:nom, ~~unum~~:pl]   VP
<pronoun[num:pl, ~~ucase~~:nom]>   [bark]
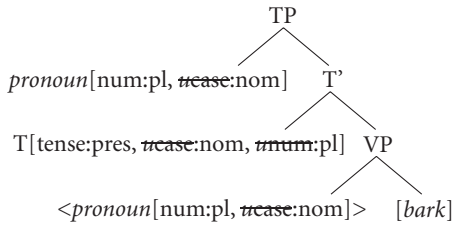
f. Associate morphemes with these feature bundles. For example, the morpheme *they* will be the spell-out of *pronoun*[num:pl, *u*case: nom], whereas, if this lexical item had had its case feature valued as [accusative], the appropriate spell-out would have been *them*. The copy of the pronoun is not pronounced.

TP
they   T'
0[+affix]   VP
bark
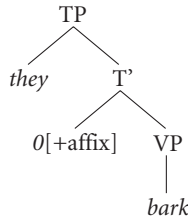
g. Perform whatever morphological operations are triggered by the featural properties of lexical items. The most important one here is the rule that realizes the inflectional features on T as a suffix on V. This rule is a morphological adjustment rule, rather than a syntactic movement rule (we show it on a tree structure here for simplicity as leaving behind an empty category notated e; we are not committed to this operation being a tree-theoretic rather than a string-theoretic operation. See Embick & Noyer 2002; Bobaljik 1995 for discussion):

TP
they   T'
e   VP
bark+0[+affix]
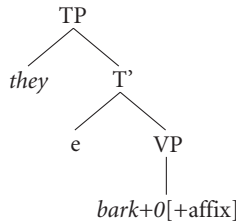
h. Perform phonological operations to derive the surface phonetic representation of (1).

Although the penultimate step makes no difference to the phonological output in this particular case, if the featural specification of T had been singular or past, T on V would have been spelled out as the (appropriate allomorphs of the) -s or -d morphemes.

This derivation maps from an initial selection of lexical items to the final output using three syntactic operations (Merge, Agree and Move) and a set of syntax/phonology interface rules (usually called the *Spell-out* component), which insert morphemes for feature bundles and may effect some reordering.

Notice that the representation which serves as input for the insertion of morphemes consists of two types of features: semantically interpretable features and checked uninterpretable features. Both types of feature can affect the phonological form, but only the first can affect the meaning. Within the spell-out component itself, we have another kind of feature which can affect phonological form (in this case, [+affix]). Since this feature is outside the syntax proper, it also cannot affect semantic interpretation.

The essential intuition that we will pursue in this paper is that variation arises from lexical items having, by the end of the syntactic derivation, the same interpretable feature specification coupled with different uninterpretable and phonological specifications. We will look at two particular cases: where lexical items bear the same interpretable but different uninterpretable features, and where inserted morphemes trigger different morphophonological processes. Both cases localise variability in the output within properties of lexical elements (feature bundles, or morphemes).

Particularly interesting challenges for this model arise in cases where we find the same features being involved in both variable and categorical patterns, and we will argue that this is a result of lexical items essentially underspecifying their syntactic requirements. We make the argument through an analysis of a number of morpho-syntactic variables in a corpus of vernacular Scots, detailed below.

3.   Morphosyntactic variation in Buckie

Buckie is a small fishing town situated on the coast 60 miles north of Aberdeen in Scotland, shown in Figure 1.

It is quite isolated in both geographic and economic terms and therefore remains relatively immune to more mainstream developments. As with similarly isolated communities (e.g. Schilling-Estes & Wolfram 1994; Poplack & Tagliamonte 1991), this is reflected in the linguistic behaviour of the community (Smith 2000). The data was collected using standard sociolinguistic methodology (Labov 1984) and is highly vernacular in nature. The data amounts to approximately forty hours of tape-recorded casual conversations which have been fully transcribed and consist of over 300,000 words. The

Figure 1. Map showing location of Buckie

speakers in the sample were born and raised in the community, and indeed the majority come from families who have been in the town for generations. They are working-class and exhibit networks that were generally confined to the community in question. The speaker sample is shown in Table 1 (see further Smith 2000).

A large number of non-standard morpho-syntactic variables exist in the dialect alongside their standard counterparts. In this paper, we concentrate on just two: *was/were* alternation, as in (3), and *do* absence in negative declaratives, as in (4):[4]

(3) a. He says 'I thocht you *were* a diver or somethin' (7:262.41)[5]
         He said 'I thought you were a diver or something.'
     b. 'Aye, I thocht you *was* a scuba diver' (7:259.21)
         'Yes, I thought you were a scuba diver.'

(4) a. She's in the huff if I *dinna* let her (g:659.13)
         She's in a bad mood if I don't allow her.
     b. God, I ø *na* ken far my ain face is here (a:654.18)
         God, I don't know where my own face is here.

Table 1. Speaker sample

Age range	Male	Female
22–31	8	8
50–60	7	7
80+	4	5

Table 2. Overall distribution of *was/were* alternation

Were	*was*	Total
46	54	1313

However, the mere presence of non-standard forms tells us little about their distribution of use across community and linguistic structure. Beginning with Labov in the 1960s (e.g. Labov 1966a), the variationist paradigm has demonstrated correlations between internal linguistic and external social factors in accounting for the observed variability. We now turn to a quantitative analysis of the data in order to establish 'why, where and when it was used, as well as by whom' (Poplack 1993:252).

3.1 *Was/were* alternation

We begin with *was/were* variation, as in (3), which is not only one of the most common features of vernacular dialects worldwide (e.g. Chambers 1995), but it is also the most widely-studied (e.g. Cheshire 1982; Cornips & Corrigan 2005).

Table 2 shows that in the Buckie sample, *was* appears in contexts of standard *were* 54% of the time. Moreover, this percentage is the product of *intra-speaker* variability: of the 39 speakers in the sample, all showed variable use.

In many dialects, there is widespread use of *was* in all contexts where standard English uses *were*, i.e. with subjects which are: 2nd person singular and plural (*you*), 1st person plural (*we*), 3rd person plural (*they*), plural NPs and existential constructions with *there* and a post-copular plural NP (e.g. Cheshire 1982). This has been claimed to be the result of analogy with other verbs that show no singular/plural distinction (Feagin 1979) or that *was* in *were* is a primitive of vernacular dialects (Chambers 1995:242).

However, when the Buckie data are divided by grammatical person, as in (5)–(9), a distinct pattern emerges: variable use in all contexts of standard *were* except with the pronoun *they*. The results are shown in Table 3.

Table 3. Distribution of *was* in *were* by grammatical person

	%	N
2nd singular *you*	69	161
1st plural *we*	67	368
3rd plural pronoun *they*	0	435
Existential *there*	90	162
NP plural	56	187

2nd person singular pronoun *you*

 (5) a. He says 'I thocht *you were* a diver or somethin' (7:262.41)
     He said 'I thought you were a diver or something.'
   b. 'Aye, I thocht *you was* a scuba diver' (7:259.21)
     'Yes, I thought you were a scuba diver.'

1st person plural pronoun *we*

 (6) a. There was one nicht *we were* lyin' at anchor (g:875.32)
     There was one night we were lying at anchor.
   b. We played on 'at beach 'til *we was* tired, sailin' boaties, bilin' whelks…
     (b:254.15)
     We played on that beach until we were tired, sailing boats, boiling
     whelks.

3rd person plural pronoun *they*

 (7) a. *They were* aie sort o' pickin' on me, like (j:504.3)
     They were always sort of picking on me.
   b. *They were* still like partying hard (j:635.28)
     They were still partying hard.

3rd person plural *full NP*

 (8) a. Buckie *boats were* a' bonny graint (g:1066.0)
     Buckie boats were all nicely grained.
   b. The *mothers was* roaring at ye comin' in (b:256.34)
     The mothers were shouting at you to come in.

Plural existential *there*

 (9) a. *There were* a puckle thatched houses like that (c:335.17)
     There were a couple of thatched houses like that.
   b. Oh, *there was* a lot of coopers 'at time (c:13.45)
     There were a lot of barrel makers at that time.
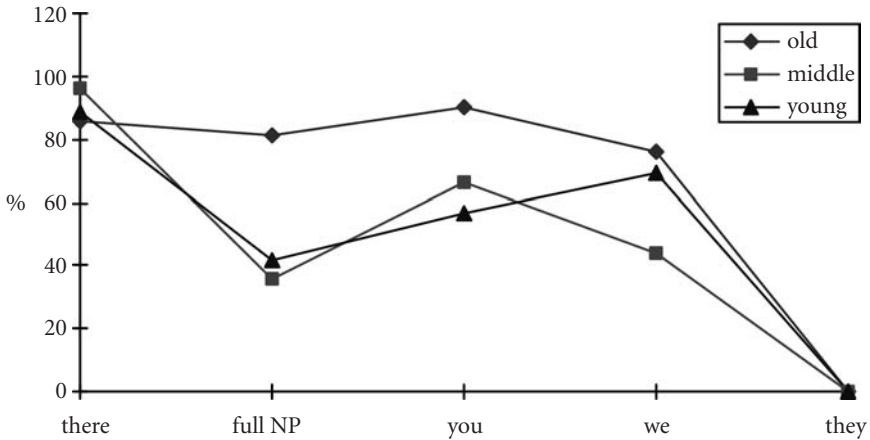
Figure 2. Distribution of *was* in *were* by grammatical person and age

When the speaker data are divided by age, there is a slight decrease in use of the non-standard form across all variable contexts from old to young, as in Figure 2.

This suggests that prescriptive norms may have an impact on this variable, but the categorical versus variable distinction remains.[6]

This pattern is, in fact, consistent with the historical record for northern dialects (e.g. Murray 1873), where an NP/pronoun distinction exists in 3rd person plural contexts: *were* is used with pronominal *they* but *was* with full NPs. This is in contrast to other dialects, which tend to show variable use across all grammatical persons (e.g. Cheshire 1982) and therefore cannot in this case be the result of 'primitive tendencies' nor analogical change as previously proposed. We return to this categorical vs. variable use in the Buckie dialect in Section 5.

3.2 *Do* absence in negative declaratives

The second variable under investigation is *do* absence, as in (4). In contrast to *was/were* alternation, this variable is not in widespread use in other varieties, and in fact appears to be restricted to rural areas on the north east coast of Scotland. Moreover, the variable context is highly circumscribed in the Buckie dialect -*do* absence only occurs in negative declarative sentences in the simple present tense. Thus, past tense negatives are always marked with *did.* However, it is similar to *was/were* alternation in that (i) all speakers exhibit variable use,

and (ii) grammatical person, as in (10)–(16) plays a major role in governing the variability.

1st person singular *I*

(10) a. She's in the huff if *I dinna* let her (g:659.13)
          She's in a bad mood if I don't allow her.
      b. God, *I ø na* ken far my ain face is here (a:654.18)
          God, I don't know where my own face is here.

2nd person singular *you*

(11) a. *Ye dinna* think ye'll be drunk (n:349.56)
          You don't think you'll be drunk.
      b. *Ye ø na* hear o' him onywye, ken (u:54.86)
          You don't hear of him anywhere, you know.

3rd person singular *he/she/it*

(12) a. *He disna* get word fae the loon (c:526.19)
          He doesn't hear any news from the boy.
      b. *It disna* cost nothin' to walk ower the hill (l:604.21)
          It doesn't cost anything to walk over the hill.

3rd person singular *full NP*

(13) a. No, *Willy disna* play much golf (@:455.56)
          No, Willy doesn't play much golf.
      b. *The car disna* ging in the garage (x:58.0)
          The car doesn't go in the garage.

1st person plural *we*

(14) a. *We dinna* really socialise that much (k:329.53)
          We don't really socialise that much.
      b. *We ø na* hae raffles (*:32.30)
          We don't have raffles.

3rd person plural *they*

(15) a. *They dinna* gie them great pay, like (4:493.26)
          They don't give them good pay.
      b. *They ø na* lose trade (*:44.32)
          They don't lose trade.

Table 4. Distribution of *do* absence by grammatical person

	%	N
1st person singular *I*	63	460
2nd person singular *you*	13	86
3rd person singular *he/she/it*	0	120
3rd person singular – full NP	0	22
1st person plural *we*	19	16
3rd person plural *they*	5	40
3rd person plural – full NP	0	12

3rd person plural *full NP*

(16) a. *Bairns dinna* coont (u:492.20)
          Children don't count.
      b. *A lot of families disna* get what that cats get (e:478.28)
          A lot of families don't get what that cat gets.

Table 4 shows the distribution of *do* absence when the data are divided in this way.

As with *was/were*, a categorical versus variable distinction arises, although the constraints are different in this case: negative *do* is variable in all contexts except for 3rd person singular pronouns and NPs, and plural NPs. However, grammaticality judgment tests show that speakers accept *do* absence with plural NPs, thus the categorical result in the data set collected for this study is likely to be the result of the small number of contexts of use (N = 12). We return to this point in Section 5.

Figure 3 shows the use of this variable across the three generations.

Unlike the patterns for *was/were* alternation, there is no difference in rates of use or constraints across three generations, thus no change in progress. Moreover, there are no gender differences: on the basis that this external factor has been demonstrated to be associated with stigmatization elsewhere (e.g. Trudgill 1972), we conclude that *do* absence is unlikely to be stigmatized in this community (see also Smith 2001a).
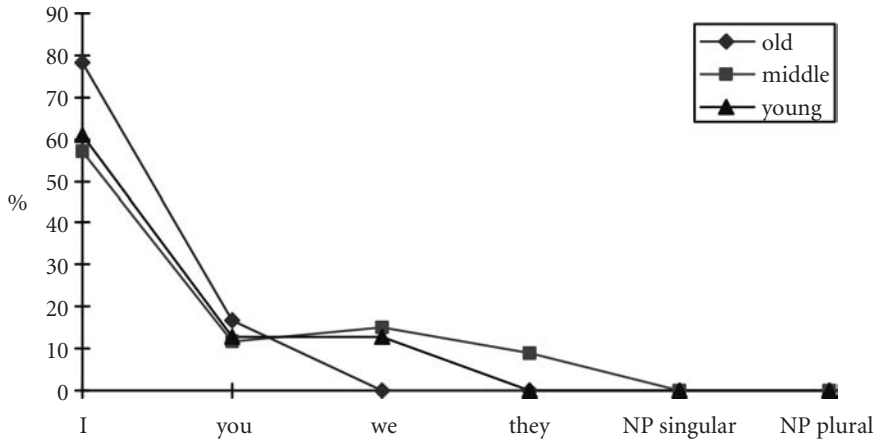
Figure 3. Distribution of *do* absence by grammatical person and age

3.3 Summary of findings

In sum, the Buckie data provides us with a number of distinct patterns for both variable features:

1. Categorical vs. variable use.
2. The particular patterns are geographically circumscribed.
3. *Was/were* shows slow change in progress towards the standard form, perhaps due to the effect of standard norms. *Do* absence is stable in the community, despite its non-standard status.

Two questions arise from these data:

1. How can a formal, universally constrained system of grammar account for the possibility of this kind of variation?

And more specifically,

2. How can such a system account for the differential patterns of categoricity versus variability as demonstrated in these data?

4. Linking syntax and variation

The examples we have discussed so far are particularly interesting in that they involve a paradigm of cases, one or more of which display variable behaviour,

with the remainder displaying categorical behaviour. The conditioning of whether the behaviour is variable or categorical in this dialect can be related to morphosyntactic features rather than sociolinguistic factors. The conditioning of the variable cases themselves has to do with both morphosyntactic and sociolinguistic factors.

In this section we propose that what lies at the core of this kind of morphosyntactically conditioned alternation between categorical and variable patterns is that the syntactic system gives the same semantic output with two distinct syntactic inputs. This means that we essentially have different syntactic representations for the variants, but those representations map to exactly the same interpretations. This is, of course, the classical definition of a linguistic variable, where given linguistic 'functions' may be realized in different forms (e.g. Labov 1966b). However, what we're attempting to do here is explicate how these two semantically equivalent forms are derived from the syntactic system in a principled way.

We implement this basic idea using ideas from the Minimalist Program as discussed in Section 2. The reason for doing this is that this framework offers us a new approach for dealing with the relationship between syntactic elements that receive a semantic interpretation and those that do not.

To see how the idea works schematically, imagine that we have two lexical items, one with the feature specification [F:a] and the other with the specification [F:a, uG:], where F is interpretable, and uG is not. If we construct a derivation using [F:a] and some lexical item with the specification [G:b], the final representation will simply look as follows:

(17)   … [F:a] … [G:b] …

If, on the other hand, we had elected to use [F:a, uG:] rather than [F:a], and we pair this with [G:b], then the final representation would look like (18), where G has checked uG:

(18)   … [F:a, ~~uG~~:b] … [G:b] …

Now notice that both (17) and (18) contain exactly the same interpretable features, and it follows that they will have exactly the same semantic interpretation. However, they are distinctive in that the feature uG is present in (18) and hence (18)'s spell-out may differ from that of (17), giving the possibility of variants.[7]

Our claim is that this basic idea is what lies at the heart of morphosyntactic variation. The dialect in question contains lexical items which differ only in whether they bear an extra uninterpretable feature or not.[8] If a lexi-

cal item bears such a feature, then the derivation must check this feature, and the surface output may be sensitive to this.

The question of the limits of morphosyntactic variation then becomes a question about the possible collocations of morphosyntactic properties of lexical items, especially functional categories. Presumably some such combinations are ruled out as part of the specification of UG, while others may be possible, but unlikely for functional or historical reasons. If this is correct, then theories of UG and theories of linguistic variation have much to contribute to each other.

Our framework contrasts with the two other major proposals for the interaction between syntactic theory and variationist theory: variable rules on the one hand, and multiple grammars/intra-language parameter setting on the other. We briefly outline these other approaches here, and then point out the differences between them and our approach.

One early attempt to connect generative linguistic theory with the idea of structured variation was the development of variable rules by Labov and his co-workers (e.g. Sankoff & Labov 1979). A variable rule essentially involves the specification of contexts for linguistic rules, where these contexts are associated with a probability index. For example, Cedergren and Sankoff (1974) report the case of variable deletion of the complementizer *que* in Montréal French. They propose the following rule:

(19)   que → <0> / <[+sib], [+cns, –sib], [–cns]> ## ___ ## <[+sib], [+cns, –sib], [–cns]>

This rule essentially says that *que* may be realized as zero in a context where the preceding word has a certain specification, and/or the following word has a certain specification. The observed variability found by the researchers was that preceding and following sibilants freely permitted deletion of *que*, but the absence of the features [+sib] (sibilant) and [+cns] (consonant) restricted the rule application. The rule is associated with probabilities in the fashion specified in Table 5.

The idea is that these probabilities come into play in applying the rule, and account for the observed distribution of the variation in terms of frequencies (which are derivative of the probability of applying the rule). The core intuition here is that variability is deeply embedded in the grammatical competence of speakers of the language as the probability index associated with the rule contexts.

An alternative approach is adopted by researchers like Bickerton (1971); De Camp (1971) and, more recently, Kroch and his co-workers (e.g. Kroch

Table 5. Constraints on variable deletion of the complementizer *que* in Montréal French

Preceding Environment	[+sib] __	[+cns, –sib]__	[–cns] __
Effect	1	.85	.37
Following Environment	[+sib] __	[+cns, –sib]__	[–cns] __
Effect	1	.50	.10
Occupational Class	Workers	Professionals	
Effect	1	.35	
Sex	Women	Men	
Effect	1	1	

1989a, 1989b; Taylor 1994; Pintzuk 1999; Santorini 1989). The idea here is that variation in social, situational and temporal domains arises from the existence of more than one grammar which speakers choose from. Rather than these grammars incorporating variability within the rule system, the variability arises from the choice of the particular grammar. Bickerton proposed such a system to account for synchronic variation, while Kroch and his associates have used this kind of approach to account for patterns of syntactic change in the history of English and other languages.

Related to this idea, but distinct from it, is recent work on dialect variation by Alison Henry (Henry 1995, 1999; Henry et al. 1997). Henry co-opts the notion of parameter, usually used in syntactic theory to account for variation between languages, and uses it instead to explain dialectal variation within what is usually considered to be a single language.[9] For example, in Belfast English, the following two examples are acceptable variants:

(20)   You go away

(21)   Go you away

Henry shows that, for some speakers, the inverted form in (21) may occur with all verbs, and argues that this is because the verb optionally moves to C in imperatives. For other speakers, the inversion is only possible with unaccusative verbs, and she proposes that in this case the subject is structurally lower (it is essentially in its VP internal base position). Henry notes that the speakers who allow inversion with all verbs are largely older speakers, and argues that this sociolinguistic fact suggests that this grammar is being lost. She proposes that this loss is due to the fact that movement of main verbs to C in English is generally disallowed, and hence the specific case of movement here is an exception and therefore dispreferred. The important point, though, is that the two dialects have a parametric difference which is embedded in different rule systems.

Both the competing grammars and the parametric approach have in common the idea that there is more than one system of grammatical knowledge in the head of the native speaker, and variation then boils down to decisions that the speaker makes about which grammatical output to choose. Of course these decisions may be below the level of consciousness (e.g. Labov 1994:78). The variable rule approach contrasts with this in that it states the variation as part of the rule, tying it in much more directly to grammatical competence.

The minimalist system we have proposed above is more akin to the competing grammars/parametric variation approaches, since there is no notion of a probability tied to a particular rule (in fact the only rules, Merge, Move and Agree, are invariant and apply categorically in particular cases). In our proposal, like that of the competing grammars/parametric variation frameworks, the notion of choice of variant is not assumed to be part of the specification of the syntactic system itself, rather it is a separate mechanism that interacts with the syntax. However, our perspective differs from these approaches in that it assumes only one invariant grammatical system, containing universal mechanisms, rather than a range of systems. Each speaker, however, has a lexicon, a memorized store of pairings of syntactic features and lexical meanings, and it is the choice of lexical item that is the source of variation. This choice is influenced by various factors: ease of lexical access (perhaps related to how common the word is), questions of speaker-hearer relationships, notions of social identity, ease of processing etc. These factors can of course be modelled as probabilities attached to the lexical entry, should one want to do so.[10] On our view, then, the mechanism that allows variation is primarily at the level of the individual, while the factors that influence variation may be either at the individual level (ease of processing etc.) or at the community level.

Notice that this is a very minimal theory, since the idea that speakers have to choose lexical items is one which we simply cannot do without. Localizing morphosyntactic variation in choice of lexical items means that we do not have to posit any special mechanism to deal with variation: variation is precisely what we should expect.[11]

5.   Analysis

In this section we provide concrete analyses of the kinds of variation we introduced in Section 2. Our purpose here is not so much to show that these are the right analyses, but rather to show that the theoretical system we adopt makes available a fruitful approach to this kind of variation.

5.1 *Was/Were* alternation

The first type of variation we will consider within this framework is the variation between *was* and *were*. Recall that with plural subjects, only the pronoun *they* categorically triggers the appearance of *were*; first and second person pronouns, and full NPs appear to both.

We adopt a standard view of the syntax of finite *be* in English: it raises and adjoins to T (e.g. Pollock 1989). In the case of auxiliary *be* we assume that it originates in some auxiliary position above the verb phrase, while in the case of copular *be*, we assume it originates inside VP. Nothing turns on these assumptions; what is important is that *be* raises to finite T if there are no other auxiliaries intervening:

(22)

```
                    T
                 ___|___
                SU      T'
                      __|___
                     T      AUXP
                  ___|___   __|___
                 be    T <be>    VP
                             ___|___
                           <SU>    V'
                                 ___|___
                                V      OBJ
```
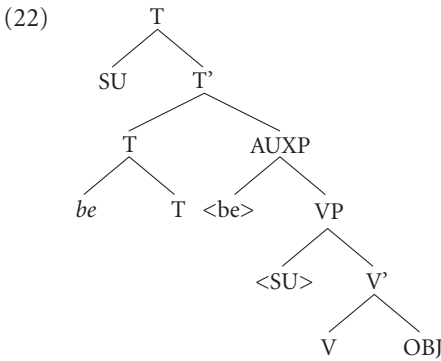
We will assume that pronouns bear interpretable person and number features. We distinguish first and second person features from third by the specifications [pers:+] for the former and [pers:–] for the latter, assuming that third person is lack of a positive specification for person (see, for example, Harley & Ritter 2002 and references therein). We will assume here that [pers:1] and [pers:2] are the two possible positive specifications for person.[12]

With this in mind, T[past] will bear unvalued features for number and person as follows:

(23) T[tense:past, *u*num:, *u*pers:]

Now, when T Agrees with a pronominal subject, these features are checked and valued. For example, we will have:

(24) T[tense:past, *u*case:nom, *u*num:, *u*pers:] … pronoun[num:pl, pers:1, *u*case:] →

> T[tense:past, ~~ucase~~:nom, ~~unum~~:pl, ~~upers~~:1] ... pronoun[num:pl, pers:1, ~~ucase~~:nom]

The spell-out of [*be* T] here will be *were* and the spell-out of the pronominal subject will be *we*.[13] Similarly, if we combine T with a third person singular, we have:

(25)   T[tense:past, *u*case:nom, *u*num:, *u*pers:] ... pronoun[gen:fem num:sing, pers:–, *u*case:] →
       T[tense:past, ~~ucase~~:nom, ~~unum~~:sing, ~~upers~~:–] ... pronoun[gen:fem, num:sing, pers:–, ~~ucase~~:nom]

This derivation will give the spell-outs *was* and *she*.

So far, we expect no variation; the spell-out of the verb will depend categorically on its featural content. However, variation will arise if there is another lexical item which can combine with the same pronominals to give the same output of interpretable features, but which has a different featural content in terms of uninterpretable features. The following lexical item, which we will call T2, will do the trick:

(26)   T2[tense:past, *u*case:nom, *u*pers:]

Now either T or T2 will be able to check with a first person plural pronoun. If we choose T, the derivation runs just as in (24), with the output *were*. If however, we choose T2, the person feature of T2 will be valued, and everything will check appropriately, giving exactly the same set of *interpretable* features as we saw with T:

(27)   T[tense:past, *u*case:nom, *u*pers:] ... pronoun[num:pl, pers:1, *u*case:] →
       T[tense:past, ~~ucase~~:nom, ~~upers~~:1] ... pronoun[num:pl, pers:1, ~~ucase~~:nom]

However, the featural content of [*be* T2] differs from that of [*be* T], and the morphology can be sensitive to this, spelling out the former as *was*. More specifically, let us propose the following spell-out for this T:

(28)   [*be* T2[tense:past, ~~ucase~~:nom, ~~upers~~:+]] spells out as *was*

We therefore have variable *were* and *was* with first person plural. Note that we have required T2 to have a positive specification for person. We will see immediately below that this is what derives the categoricality of third person plural *they were*.

What about the other cases? First person singular will also be able to check with both lexical entries, but will give *was* as output in either case, since [1] counts as a positive value for [*u*pers]:

(29) a. [*be* T[tense:past, ~~*u*case~~:nom, ~~*u*num~~:sing, ~~*u*pers~~:1]] spells out as *was*
      b. [*be* T2[tense:past, ~~*u*case~~:nom, ~~*u*pers~~:1]] spells out as *was*

Third person singular will combine with T2, but there is no spell-out for a minus version of [*u*pers] on T2 without an associated number feature:

(30) a. [*be* T[tense:past, ~~*u*case~~:nom, ~~*u*num~~:sing, ~~*u*pers~~:–]] spells out as *was*
      b. [*be* T2[tense:past, ~~*u*case~~:nom, ~~*u*pers~~:–]] no spell-out!

Second person works just like first person plural, combining with either T to give a variable output:

(31) a. [*be* T[tense:past, ~~*u*case~~:nom, ~~*u*num~~:pl, ~~*u*pers~~:2]] spells out as *were*
      b. [*be* T2[tense:past, ~~*u*case~~:nom, ~~*u*pers~~:2]] spells out as *was*

Third person plural, which was categorical, cannot value [*u*pers] on T2 as positive, and so there is no appropriate spell-out. It follows that there is no derivation leading to the output *they was*.

(32) a. [*be* T[tense:past, ~~*u*case~~:nom, ~~*u*num~~:pl, ~~*u*pers~~:–]] spells out as *were*
      b. [*be* T2[tense:past, ~~*u*case~~:nom, ~~*u*pers~~:–]] → no spell-out!

The essential intuition behind all of these cases of variability, is that there are two lexical items either of which can combine with some subset of the pronominal paradigm. Once the various features of these lexical items are checked, the final output is identical in terms of interpretable features, and so the meaning is the same in both cases. However, the morphological form of the spell-outs attached to these lexical items can be sensitive to all of their features, hence the surface form may vary. The variable/categorical split is due to the interaction of the features of the pronouns, and the morphological well-formedness of the features of the *be* plus T composite.

The Buckie system is relatively rare in having this variable/categorical split. Many other systems allow variability throughout the paradigm (e.g. Cheshire 1982). This is straightforwardly captured by assuming that [*be* T2] in these other systems spells out as *was* irrespective of the value of [*u*pers].

5.2 Variability in NP agreement

Finally we turn to *was/were* variability with plural NPs, which contrasts with the categoricality of plural agreement with *they*.

The split between the capabilities of full NPs and pronouns to trigger agreement is one which is well-established for other languages. For example, in Welsh, subject pronouns trigger agreement on their verbs:

(33) Gwelodd ef y   car
      Saw   he the car
      'He saw the car.'

(34) Gwelsant hwy y   car
      Saw-3PL they the car
      'They saw the car.'

However, this contrasts with full NPs, which do not trigger subject agreement:

(35) Gwelodd y   dyn y   car
      Saw   the man the car
      'The man saw the car.'

(36) Gwelodd y   dynion y   car
      Saw   the men   the car
      'The men saw the car.'

(37) *Gwelsant y   dynion y   car
      Saw-3PL   the men   the car
      'The men saw the car.'

We will assume then that UG makes available a pronoun/NP split, as well as a singular/plural split. We implement this in our feature system by assuming that full NPs consist of a determiner layer selecting an NP (Abney 1987):

(38)
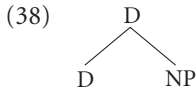
When the head noun bears an interpretable number feature, this feature may, or may not, appear on the selecting D, depending on the lexical specification of D. Welsh DPs, then, uniformly have singular determiners, and it is the determiners that trigger number agreement on the verbs:[14]

(39) V-[$u$num] [$_{DP}$ D[$u$num:sing] NP[num:pl] $\rightarrow$
      V-[~~$u$num~~:sing] [$_{DP}$ D[~~$u$num~~:sing] NP[num:pl]

There is some intriguing evidence that Buckie follows this pattern. Demonstratives display variable agreement for number with their head noun:

Plural proximate

(40) a.   I'd a' *these* cuttings (j:987.1)
           I had all these cuttings.
      b.   My mam had all *this* stories o' ootside folk (m:903.4)
           My mother had all these stories about foreign people.

Plural distal

(41) a.   She's one of *those* bonny big houses (6:311.1)
           She has one of those lovely big houses.
      b.   But I 'd piles of *that* photos of the dancing (e:212.8)
           But I had piles of those photographs of the dancing.
      c.   That was ain o' *them* grogain suits (a:971.4)
           That was one of those grogain suits.
      d.   It was a' bonny, able drifters, *thon* steel drifters
           They were all good able drifters, those steel drifters.

Singular distal

(42) a.   Did he mairry again after Linda gied away and marriet *thon* loon?
           (j:675.42)
           Did he marry again after Linda went away and married that guy?
      b.   We 'd forty staff employed in *that* job (*:98.7)
           We had forty staff employed in that job.

Here we have a number of variants used, including the relic form *thon/yon* which is still used in some Celtic English dialects and the pandialectal use of *them*.

Table 6 shows the distribution of forms across these different contexts of use.

In singular contexts, there is virtually no variation: in this case, the standard form *that* is used 97% of the time and the relic form *thon* only 3%.

In contrast, plural demonstratives demonstrate high rates of use of non-agreement in both proximate (66% *this*) and distal contexts (69% *that*). In addition, there are relatively high rates of the older form *thon* in this context.

We will not provide an analysis of this pattern here, for reasons of space, but merely assume that it provides good evidence that D in Buckie may be specified as singular. On our account this will be a property of the lexical items which have the categorial feature D.

Table 6. Distribution of demonstratives in different contexts of use

	N	%
Distribution of *singular distal demonstratives*		
That	1991	97
Thon	55	3
Distribution of *plural proximate demonstratives*		
These	33	34
This	65	66
Distribution of *plural distal demonstratives*		
Those	7	4
That	132	69
Thon	40	21
Them	13	7

Given this, we now have an explanation for *was/were* variability with plural DPs. Recall that T is specified as follows (T2 is irrelevant here since DPs are [pers:–]):

(43)   T[tense:past, *u*case:nom, *u*num:, *u*pers:]

If the subject is a DP, it will variably have the specification [*u*num:sing]/ [*u*num:pl] on D, perhaps depending on whether the feature is lexically specified as valued or not:

(44)   T[tense:past, *u*case:nom, *u*num:, *u*pers:] … DP[*u*num:sing, pers:–, *u*case:] →
       T[tense:past, ~~*u*case~~:nom, ~~*u*num~~:sing, ~~*u*pers~~:–] … DP[~~*u*num~~:sing, pers:–, ~~*u*case~~:nom]
       The mothers was …

(45)   T[tense:past, *u*case:nom, *u*num:, *u*pers:] … DP[*u*num:pl, pers:–, *u*case:] →
       T[tense:past, ~~*u*case~~:nom, ~~*u*num~~:pl, ~~*u*pers~~:–] … DP[~~*u*num~~:pl, pers:–, ~~*u*case~~:nom]
       The mothers were …

The variability we see here, then, depends not on multiple lexical entries for T, but rather on multiple lexical entries for D.

5.3 *Do*-absence

The phenomenon of *do*-absence, discussed in Section 3, is variable in contexts of non-third person singular agreement. It is categorical in past tense con-

texts, and with third person singular subjects, irrespective of whether they are pronominals or full NPs.

The analysis we propose for this phenomenon also localises the possibility of variation in the properties of lexical elements. However, what we see here does not derive from choice of a lexical item *qua* feature bundle, but rather from the choice of morpheme associated with a lexical item by the spell-out mechanisms.

As discussed in Section 2, we assume that the operation that inflects the main verb is a morphological operation that attaches finite T to the verb as a suffix. As is well known, this operation is sensitive to whether negation intervenes between T and the verb. If negation does intervene, then the operation cannot take place, and a dummy verb *do* is inserted to undergo inflection:

(46)   The books -ed[+affix] inform us all
       → The books inform-ed[+affix] us all

(47)   The books -ed[+affix] not inform us all
       → The books *do*-ed[+affix] not inform us all

The morphological interpretation of V+T may be far more complex than simple concatenation (involving, for example, ablaut, zero-realisation etc.), and we will not take a stand on how such morphology is accomplished (whether by readjustment rules, paradigmatic look-up, rules of morphological referral, etc.).

In the Standard English present tense, except for third person singular, the morpheme associated with T is null, but it must still be assumed to be an affix, since we still find *do*-support:

(48)   The books -0[+affix] inform us all
       → The books inform-0[+affix] us all

(49)   The books -0[+affix] not inform us all
       → The books do-0[+affix] not inform us all

We can straightforwardly capture the variation in *do*-absence found in the Buckie dialect by assuming that the zero morpheme associated with non-third singular T simply has two forms: one is [+affix] while the other is not. This kind of surface variation is well-studied in variationist work: for example, the difference between *in/ing* variants in different varieties of English (e.g. Houston 1991). We therefore have the two following options:

(50)   T[tense:pres, *u*pers:–, *u*num:sing] → -z[+affix]

(51)   T[tense:pres] → -0[+affix]; 0[–affix]

To see this more explicitly, take an example like (52) below:

(52)   [_TP pronoun[pers:–, num:pl, ~~ucase~~:nom] T[tense:pres, ~~upers~~:–,
        ~~unum~~:pl, ~~ucase~~:nom] Negation [V NP]]

We now need to choose a morpheme for T. Clearly we cannot choose the morpheme -z[+affix], since it does not match T's features. We therefore insert the elsewhere morpheme for present tense which has two allomorphs. This gives us:

(53)   They -0[+affix] na lose ...

(54)   They 0[–affix] na lose ...

Of these, only the former triggers *do*-support:

(55)   They do-0[+affix] na lose ...

(56)   They 0[–affix] na ken ...

The variation we see here, then, reduces to choice of allomorph for a particular morpheme. Essentially we have standard allomorphy but without a conditioning context. This is different from the variation we saw in the preceding section, which arose from the different lexical items (feature bundles) that entered the derivation, but it is still crucially localised in properties of lexical elements.

Finally, we can account for the variability in the ability of full DPs to license *do*-absence as a result of the idea, motivated earlier, that DPs containing plural nouns need not bear plural agreement features at the D level. Consider a derivation where the DP *the quines* ('the girls') is the subject. Prior to movement of this subject to the specifier of TP, we have:

(57)   T [*u*num:]... [_DP the[*u*num:pl] quines[num:pl]]

When T Agrees with D, its uninterpretable feature will be valued and checked, and the spell-out associated with this feature bundle [*u*num:pl] is zero, with its two variants. This will predict the eventual forms:

(58)   a.   The quines do na ken
              The girls do not know.
        b.   The quines na ken
              The girls do not know.

If D however bears the feature [*u*num:sing], then T will match and value as T[*u*num:sing]. T in this case has the /-z/ spell-out, and forces *do*-support. We therefore correctly predict the (categorical) contrast between (55) and (56):

(59)   The quines doesna ken
        The girls do not know.

(60)   *The quines na kens

Once again, note that the interpretable features of (58) and (59) are identical: all three variants are predicted to have the same meaning.

6.   Conclusion

Our basic proposal here has been that a Minimalist approach to syntax melds extremely well with the kinds of data that variationists study because it has two core properties: it builds the notion of (tacit) choice of lexical item into the syntactic system, and it allows derivations with different lexical items to converge on the same basic semantic representation, thus capturing the multiple form/single meaning notion of a linguistic variable. The orderly patterns of variation seen across (groups of) individuals reduces to the lexical choice by an individual speaker of functional elements with particular feature specifications. This choice is influenced by a range of use-related factors, such as processing, frequency of individual lexical items in the register or community discourse, and broader sociolinguistic and communicative factors. These use-related influences are, for us, outside the grammar proper, which is simply a specification of the syntactic, semantic and morphophonological properties of lexical items together with an invariant syntactic engine sensitive to these properties. They are, however, part of language in the broader sense, and their study may impact on our understanding of how I-language is embedded within other cognitive mechanisms.

We would like to emphasize that our purpose in this paper has not been to provide a theory of how these various factors influence lexical choice, but rather to highlight the usefulness of drawing a distinction between (i) the mechanism which allows variability in an essentially invariant (minimalist) syntactic system, part of I-language, and (ii) factors which may be related to individuals' biologically constrained capacities to use language (e.g. processing, prosodic or information structure theoretic factors) or to the (possibly tacit) desire of individuals to conform to, or to rebel against, their communities' impositions.

Notes

* We would like to thank the participants of the ICLAVE workshop on syntax and variation, Uppsala, 2003, and Annabel Cormack, Miriam Meyerhoff, Susan Pintzuk, Christer Platzack, Bernadette Plunkett, Ann Taylor, Graeme Trousdale, George Tsoulas, Tonjes Veenstra and Anthony Warner for helpful comments. We would also like to thank Leonie Cornips, Karen Corrigan and two anonymous reviewers for extensive feedback on this paper.

1. We use the mechanisms of uninterpretable features, valuation and deletion here, but we are not committed to this implementation. See Chomsky (2002) for a system which reduces uninterpretability to valuation, and Adger and Ramchand (to appear) for a system which further removes the need for a deletion operation.

2. There are alternative methods of dealing with the spell-out of the inflectional features on V in English, some purely syntactic, some as interface rules operating between syntactic and morphological structures (see Adger 2003 for an example of the former approach); we adopt something close to the standard here for simplicity and familiarity. We have glossed over a number of questions about whether there is a more articulated structure above the VP (see, for example, Pollock 1989 and much subsequent work), within the VP (see, for example, Larson 1987; Hale & Keyser 1993, among many others), and have assumed a very simple analysis of case checking. None of these simplifications affect the material point.

3. These lexical items are simplified. We follow Distributed Morphology (Halle & Marantz 1993) in assuming that they lack any phonological information: they are just bundles of syntactic and semantic features which are spelled out as morphemes at some point in the derivation.

4. See Adger and Smith (2002) for an application of the ideas discussed here to the structure and use of negative concord in the same speech community.

5. Codes represent speaker and place in the transcription.

6. A number of other constraints were tested in the data, including polarity and copula vs. auxiliary status of the verb. However, none of these were significant for the use of non-standard *was*.

7. Note that by spell-out here we actually mean more than just the pronunciation of the feature; uninterpretable features play a crucial role in syntactic movement as well, so that variability in word order and correlations between morphology and word order can also be handled by this system (see Adger & Smith 2002 for discussion of how this works in negative concord constructions).

8. One concern that may come to mind here is blocking effects. It is often assumed that the lexicon is structured so that it does not contain items with identical properties. From our perspective, the presence of an uninterpretable feature on one lexical item is enough to distinguish it from another, and hence this sort of blocking effect does not come into play.

9. See Corrigan (1997) for an alternative perspective on variation in Irish English.

10. It is not clear to us that this is the correct way to deal with all constraints on variability, although it is clear that language learners are sensitive to statistical patterns in variable input

(e.g. Saffran, Aslin, & Newport 1998). Whether the processor is also sensitive to probabilities is a controversial matter.

11. A referee points out that it is not clear how variation such as the Jespersen cycle of negation, or verbal clusters in West Germanic can be related to the idea that all morphosyntactic variation is essentially lexical. On negation, see Adger and Smith (2002), where we argue that variation in negative concord arises from the interaction between movement of verbs and negative XPs, both lexically triggered. We will not speculate on verbal cluster variation here.

12. Questions arise here about implementation, since what we have just stipulated does not follow from the formalism we have adopted. See, for example, Harbour (2003), or Harley and Ritter (2002), for different ways of deriving these results from a theory of phi-features.

13. We assume here that morphemes are spell-outs of fully specified feature bundles and abstract away from the question of how partial specification of morphemes affects lexical insertion, as in theories like Distributed Morphology (see Halle & Marantz 1993 for discussion).

14. We assume here that the number feature on D is uninterpretable, and that interpretable number is carried by the lexical category lower down, or by some functional element (see, for example, Ritter 1991; Borer 2004).

References

Abney, Steven (1987). The English Noun Phrase in its Sentential Aspect. Ph.D. dissertation, MIT.

Adger, David (2003). *Core Syntax: A Minimalist Approach*. Oxford: Oxford University Press.

Adger, David & Gillian Ramchand (to appear). "Merge vs. Move: Wh-Dependencies Revisited". *Linguistic Inquiry*.

Adger, David & Jennifer Smith (2002). "Negative Concord in Buckie". Paper presented at the *North East Syntax Seminar*, University of Edinburgh.

Bickerton, Derek (1971). "Inherent Variability and Variable Rules". *Foundations of Language, 7*, 457–492.

Bender, Emily (2001). Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence. Ph.D. dissertation, Stanford University.

Bobaljik, Jonathan (1995). Morphosyntax: The Syntax of Verbal Inflection. Ph.D. dissertation, MIT.

Borer, Hagit (2004). *Structuring Sense*. Oxford: Oxford University Press.

Cedergren, Henrietta & David Sankoff (1974). "Variable Rules: Performance as a Statistical Reflection of Competence". *Language, 50*, 333–355.

Chambers, Jack (1995). *Sociolinguistic Theory*. Oxford and Cambridge: Blackwell.

Cheshire, Jenny (1982). *Variation in an English Dialect*. Cambridge: Cambridge University Press.

Chomsky, Noam (1957). *Syntactic Structures*. Mouton: de Gruyter.

Chomksy, Noam (1986). *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.

Chomsky, Noam (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam (2000). "Minimalist Inquiries: The Framework". In Roger Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by Step: Essays on Minimalist Syntax in Honour of Howard Lasnik* (pp. 89–115). Cambridge, MA: MIT Press.

Chomsky, Noam (2001). "Derivation by Phase". In Michael Kenstowicz (Ed.), *Ken Hale: A Life in Language* (pp. 1–52). Cambridge, MA: MIT Press.

Chomsky, Noam (2002). *Beyond Explanatory Adequacy*. MIT Occasional Papers in Linguistics 20. Department of Linguistics and Philosophy. Cambridge, MA: MIT.

Cornips, Leonie & Karen Corrigan (2005). "Convergence and Divergence in Grammar". In Peter Auer, Frans Hinskens, & Paul Kerswill (Eds.), *Dialect Change: Convergence and Divergence in European Languages* (pp. 96–134). Cambridge: Cambridge University Press.

Corrigan, Karen (1997). "The Acquisition and Properties of a Contact Vernacular Grammar". In Anders Ahlqvist & Věra Čapkovà (Eds.), *Dán Do Oide: Essays in Memory of Conn R. Ó Cléirigh* (pp. 75–93). Dublin: The Linguistics Institute of Ireland.

De Camp, David (1971). "Towards a Generative Analysis of a Post-Creole Speech Continuum." In Dell Hymes (Ed.), *Pidginization and Creolization of Languages* (pp. 349–370). Cambridge: Cambridge University Press.

Embick, David & Rolf Noyer (2001). "Movement Operations after Syntax". *Linguistic Inquiry, 32*, 555–596.

Feagin, Crawford (1979). *Variation and Change in Alabama English: A Sociolinguistic Study of the White Community*. Washington, DC: Georgetown University Press.

Hale, Kenneth & S. Jay Keyser (1993). "On Argument Structure and the Lexical Expression of Syntactic Relations". In Kenneth Hale & S. Jay Keyser (Eds.), *The View from Building 20* (pp. 53–110). Cambridge, MA: MIT Press.

Halle, Morris & Alec Marantz (1993). "Distributed Morphology and the Pieces of Inflection". In Kenneth Hale & S. Jay Keyser (Eds.), *The View from Building 20* (pp. 111–176). Cambridge, MA: MIT Press.

Harbour, Daniel (2003). Elements of Number Theory. Ph.D. dissertation, MIT.

Harley, Heidi & Elizabeth Ritter (2002). "Person and Number in Pronouns: A Feature Geometric Analysis". *Language, 78*, 482–526.

Henry, Alison (1995). *Belfast English and Standard English: Dialect Variation and Parameter Setting*. New York and Oxford: Oxford University Press.

Henry, Alison (1999). "Minimalist Syntax and Language Variation: Expletives and Agreement". Paper presented at *NWAV 28*, University of Toronto, Canada.

Henry Alison, R. MacLaren, J. Wilson, & C. Finlay (1997). "The Acquisition of Negative Concord in Non-Standard English". In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*. MA: Cascadilla Press.

Houston, Anne (1991). "A Grammatical Continuum for (ING)". In Peter Trudgill & Jack Chambers (Eds.), *Dialects of English: Studies in Grammatical Variation* (pp. 241–257). London: Longmans.

Kayne, Richard (1989). "Notes on English Agreement". *CIEFL Bulletin, Hyderabad, 1*, 41–67.

Kroch, Anthony (1989a). "Function and Grammar in the History of English: Periphrastic *Do*". In Ralph Fasold & Deborah Schiffrin (Eds.), *Language Change and Variation* (pp. 133–172). Washington, DC: Georgetown University Press.

Kroch, Anthony (1989b). "Reflexes of Grammar in Patterns of Language Change". *Language Variation and Change, 1*, 199–244.

Labov, William (1966a). *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, William (1966b). "The Linguistic Variable as a Structural Unit". *Washington Linguistics Review, 3*, 4–22.

Labov, William (1972a). *Sociolinguistic Patterns.* Philadelphia: University of Pennsylvania Press.

Labov, William (1972b). "Negative Attraction and Negative Concord in English Grammar". *Language, 48*, 773–818.

Labov, William (1984). "Field Methods of the Project on Linguistic Change and Variation". In John Baugh & Joel Sherzer (Eds.), *Language in Use: Readings in Sociolinguistics* (pp. 28–53). Englewood Cliffs, NJ: Prentice-Hall.

Labov, William (1994). *Principles of Linguistic Change: Internal Factors*. Cambridge, MA: Blackwell.

Larson, Richard (1987). "On the Double Object Construction". *Linguistic Inquiry, 19*, 33–91.

Lasnik, Howard (1981). "Restricting the Theory of Transformations: A Case Study". In Norbert Hornstein & David Lightfoot (Eds.), *Explanation in Linguistics* (pp. 152–173). Longman: New York.

Milroy, James & Leslie Milroy (1985). *Authority in Language: Investigating Language Prescription and Standardisation.* London and Boston: Routledge and Kegan Paul.

Murray, James A. H. (1873). *The Dialect of the Southern Counties of Scotland: Its Pronunciation, Grammar and Historical Relations.* London: Philological Society.

Pesetsky, David & Esther Torrego (2001). "T to C Movement: Causes and Consequences". In Michael Kenstowicz (Ed.), *Ken Hale, a Life in Language* (pp. 355–426). Cambridge, MA: MIT Press.

Pintzuk, Susan (1999). *Phrase Structures in Competition: Variation and Change in Old English Clause Structure*. New York: Garland

Pollock, Jean-Yves (1989). "Verb Movement, Universal Grammar and the Structure of IP". *Linguistic Inquiry, 20*, 365–424.

Poplack, Shana (1993). "Variation Theory and Language Contact". In D. Preston (Ed.), *American Dialect Research* (pp. 251–286). Amsterdam and Philadelphia: John Benjamins.

Poplack, Shana & Sali Tagliamonte (1991). "African American English in the Diaspora: The Case of Old Line Nova Scotians". *Language Variation and Change, 3*, 301–339.

Ritter, Elizabeth (1991). "Two Functional Categories in the Noun Phrase". In Susan Rothstein (Ed.), *Syntax and Semantics 26* (pp. 37–62). New York: Academic Press.

Saffran, Jenny R., Richard N. Aslin, & Elissa L. Newport (1996). "Statistical Learning by 8-Month-Olds". *Science, 274*, 1926–1928.

Sankoff, David & Wiliam Labov (1979). "On the Uses of Variable Rules". *Language and Society, 8*, 189–222.

Santorini, Beatrice (1989). "Variation and Change in Yiddish Subordinate Clause Word Order". *Language Variation and Change, 5*, 257–283.

Schilling-Estes, Natalie & Walter Wolfram (1994). "Convergent Explanation and Alternative Regularization: *Were/Weren't* Leveling in a Vernacular Variety of English". *Language Variation and Change, 6*, 273–302.

Smith, Jennifer (2000). Synchrony and Diachrony in the Evolution of English: Evidence from Scotland. Ph.D. dissertation, University of York, England.

Smith, Jennifer (2001a). "*Ye ØNa Hear That Kind o' Things:* Negative *Do* in Buckie". *English World Wide, 21*, 231–259.

Smith, Jennifer (2001b). "Negative Concord in the Old and New World: Evidence from Scotland". *Language Variation and Change, 13*, 109–134.

Taylor, Anne (1994). "Variation in Past Tense Formation in the History of English". *University of Pennsylvania Working Papers in Linguistics, 1*, 143–159.

Trudgill, Peter (1972). "Sex, Covert Prestige and Linguistic Change in Urban British English". *Language in Society, 1*, 179–195.

Principles and parameters in change*

Elly van Gelderen
Arizona State University

1.   Introduction

In this paper, I examine variation in use between pronouns and nouns, between differently case marked pronouns and nouns, and between the different persons. These differences do not constitute absolute tendencies, or splits, and they are only detectable when looking at large numbers of instances. I show how this variation has theoretical implications, in particular for Economy Principles, thus providing insights into I(nternalized)-language. I also consider historical data which show that fast change (e.g. the loss of morphological case) is suggestive of a parameter resetting, but that slow change (e.g. the noun/pronoun split) is indicative not of a change in a principle but of a change in pronominal status. Using three corpora, I also raise some questions concerning the nature of linguistic data.

1.1   Background

Some variationist work within generative grammar has revolved around invoking different parametric settings to account for different varieties (e.g. Henry 1995; Kayne 1989). I will look at nominal and pronominal variation in terms of a (Minimalist) Economy Principle that guides speakers to build the syntactic structure up to just a head rather than to a phrase. This predicts that personal pronouns are less often coordinated (since coordinates are phrasal) than nouns and this is borne out consistently in the corpora used. I call these differences 'splits', adapting a phrase from typology. Looking at these contrasts in a diachronic corpus, there is very little change in 400 years, which is to be expected if we are dealing with an (invariable) principle. The changes that do occur are

indicative of pronominal change. The loss of morphological case, in contrast, is a fast change, it being a parameter resetting rather than a principle at work.

The paper helps to show that variation is relevant to theory building (cf. also Wilson & Henry 1998; Cornips & Corrigan 2005). It also discusses where electronic texts and corpora are helpful and where they are not. The outline is as follows. In Section 1.2 immediately below, I provide some background to the use of statistics and corpora in generative grammar. In Section 2, evidence for some noun-pronoun, case, and person splits in a modern corpus is given. A theoretical account for this split is presented in Section 3. In Section 4, I compare Shakespearean English to Modern English. In Section 5, I examine a pronominal change in case marking that shows quite a fast rate of change.

1.2   The use of corpora and statistics

I will first briefly discuss the corpora and texts I use. Then, I assess some generative attitudes towards the use of statistics and corpora, and justify how numbers of occurrences do indeed say something about internalized language, also known as I-language or competence, the object of generative inquiry.

In this paper, three corpora are used (see references for URLs). The rather formal (2 million-word) *Corpus of Spoken Professional American English* (hence *CSE*) consists of three parts: White House briefings (WH), Faculty Meetings at UNC (FAC), and Committee Meetings held all around the United States to discuss tests (COM). The *British National Corpus* (hence *BNC*) consists of a spoken and written part. It is a much larger corpus than the *CSE* with the spoken part comprising 10% of the 100 million-word corpus. The *BNC* is a lot less formal than the *CSE*. However, the use of the *BNC* is sometimes not practical. For instance, in the spoken *BNC*, there are 89,390 instances of *he* and *him* and 332,315 of *I* and *me*.

The *Helsinki Corpus of English Texts* (hence *HC*) is a diachronic corpus using many different text types. Old English (OE) is usually considered to be the form of language dating between 450 and 1150; Middle English (ME) between 1150 and 1500; and Early Modern English (EMOD) after 1500. The *HC* divides each further into OE1-2 from before 950; OE3 from 950 to 1050; OE4 from 1050 to 1150; ME1 from 1150 to 1250; ME2 from 1250 to 1350; ME3 from 1350 to 1420; ME4 from 1420 to 1500; EMOD1 from 1500 to 1570; EMOD2 from 1570 to 1640; and EMOD3 from 1640 to 1710.

In addition to the corpora, I have used electronic texts (made available by the *Oxford Text Archive*): the Old English *Beowulf*, the Early Middle English versions of Layamon's *Brut*, and the 1623 *First Folio* (henceforth *F1*) edition

of Shakespeare's plays. I have also used the *Dictionary of Old English* Corpus (*DOE*), available through the University of Toronto.

Generative linguists working on 'living' languages often view work with corpora and statistics as not indicative of the I-language but rather of the E(xternalized)-language (also known as performance). Wasow (2002) reviews Chomsky's views on the use of quantitative data. Some of the quotes he chooses show that what Chomsky has in mind is word choice or word concordances, not grammatical phenomena. For instance, the choice of *Nevada* over *New York* seems irrelevant to underlying linguistic structures in:

> It seems that probabilistic considerations have nothing to do with grammar, e.g. surely it is not a matter of concern for the grammar of English that 'New York' is more probable than 'Nevada' in the context 'I come from–'.
>
> (Chomsky 1962: 128)

In this paper, I try to get at subtle grammatical variation and I find that data obtained from corpora and other electronic texts show systematic differences that are indicative of features and parameters of the internalized grammar. For instance, when coordination is more frequent with a particular word class, i.e. pronouns, it indicates an important fact about the status of pronouns in the I-language.

In contrast to synchronic linguists, historical generative linguists have embraced work with e.g. the *HC*, the *DOE*, the *Penn-Helsinki Parsed Corpus of Middle English*, and the *Brooklyn-Geneva-Amsterdam-Helsinki Parsed Corpus of Old English* (e.g. Pintzuk 1999; van Bergen 2003; Trips 2002; and Wood 2003). Notable exceptions exist of historical linguists not making use of these corpora, e.g. Lightfoot (1999). Lightfoot argues that change in language (E-language) is gradual, but change in grammar (I-language) is abrupt. The latter is due to parameter resetting. His interest in abrupt changes may be the reason for his non-use of the corpora. I uncover grammatical splits and find that data obtained from corpora and other electronic texts show systematic differences that are indicative of principles and parameters of the internalized grammar.

Principles are "language-invariant statements" (Chomsky 1995: 25) whereas parameters must be set for certain values. A possible parameter is whether *wh*-movement applies overtly (so it is visible) or covertly (so it looks as if the *wh*-element is not moved) in a particular language. Examples of principles include a 'Locality Condition' on movement, 'Full Interpretation', and an 'Economy of Derivation' Principle (see Chomsky 1995: 28). The principle I assume in Section 3 belongs under Economy of Derivation.

2.   Noun/pronoun splits

In this section, I first show that pronouns are less often coordinated than nouns. They are also less often modified by relative clauses or adjectives. For practical purposes, I focus on first and third person pronouns mostly, and on coordination rather than modification. Auxiliaries also cliticize more readily to pronouns. The reason for the lack in coordination and more frequent cliticization is given in Section 3, namely pronouns can be heads more readily.[1] Second, nominative pronouns are less often coordinated than accusative ones (even though this is only statistically significant in larger, less formal corpora such as the *BNC*, as shown in Section 4). Third, first person singular is less often coordinated as compared to third person singular. The reason behind these two facts is addressed in Section 3.

2.1   Pronouns vs. nouns

From the *Corpus of Spoken Professional American English* (as mentioned, from now on *CSE*), I have selected the 161,000-word Faculty Meetings' part, but the other parts are very similar. These transcripts show a split between nouns and pronouns. Thus, nouns such as *faculty, departments* and *school(s)* are coordinated over 10% of the cases, e.g. *faculty* occurs 353 times and is coordinated with *and,* as in (1), 62 times, i.e. 17.6%. As will become clearer later on, when I use 'noun', technically this means a D(eterminer)P(hrase):[2]

   (1)   to try to tap into what *students and faculty* have an interest in doing.
                                                                      (*CSE*-FAC97)

Faculty occurs much more often than 353 times but I have disregarded the modifying uses of *faculty*, as in (2):

   (2)   but for the grieving *faculty* member who feels that he or she was dismissed
         …   (*CSE*-FAC95)

For the noun *school(s)* in the same part of the *CSE*, the percentage coordinated is 16.4, namely 55 instances of *school(s)* with nine of these coordinated. In the same part, *students* is coordinated 51 times out of 367, which is 13.9%. *Departments* occurs 52 times of which 12 are coordinated, i.e. 23%. This use is very different from that of pronouns, as Table 1 shows. Thus, first person singular pronouns are coordinated less than 1% of the time and third person singular less than 2%.

Table 1. First and third person pronouns versus nouns in CSE-FAC

	Uncoordinated	Coordinated	Total
I/me	3024	20 (= 0.66%)	3044
s/he, her/him	227	4 (= 1.73%)	231
faculty, student(s), school(s), departments	693	134 (= 16.2%)	827

χ^2 is 418.061 p < 0.001 for first persons against nouns and 33.340 p < 0.001 for third persons against nouns.

Table 2. Cliticization to pronouns in CSE-FAC

	Uncliticized	Cliticized	Total
I	2037	685 (= 25.%)	2722
you	1176	162 (= 12.1%)	1338
he	128	19 (= 12.9%)	147

Significant between first and second as well as first and third at p < 0.001 (the respective χ^2 are 41.801 and 11.284).

A second difference between nouns and pronouns is cliticization of an auxiliary. This is common with pronouns, as Table 2 shows for cliticization of *am, will, would, has* etc. to the pronoun. It occurs quite frequently with first person pronouns, as in (3), but it never occurs with *faculty, student, department*, and *school*:

(3)   *I'm* concerned that this perception came across.   (*CSE*-FAC95)

2.2 Case and person

In order to come to a better understanding of what makes pronouns behave differently, I will examine case and person in this subsection.

In Table 1, nominative and accusative forms are not separated. If one separates the nominative pronoun from the accusative in *CSE*-FAC, as in Table 3, a difference can be observed but not one that is statistically significant. The *CSE* is a relatively formal corpus and that shows in the adherence to prescriptive case rules such as having nominatives in subject position. As a result, *me* is never used as subject in e.g. *The president and me held a press conference*, unlike in e.g. the *BNC* as shown in Section 4. This may influence the results.

Table 3. Case and coordination: first and third person pronouns in CSE-FAC

	Uncoordinated	Coordinated	Total
Nominative	2884	21 (= .73%)	2905
Accusative	367	3 (= .8%)	370

$\chi^2 = 0.035$, $p > 0.50$.

Table 4. Coordination and function of the noun 'faculty'

	Uncoordinated	Coordinated	Total
Subject	69	7 (= 9%)	76
Non-subject	222	55 (= 19.9%)	277
Total	291	62 (= 17.6%)	353

$\chi^2 = 4.667$, $p < .05$.

An example of a coordinate nominative pronoun is (4):

(4)   *Barbara and I* are very excited and optimistic about the work we are undertaking.   (*CSE*-FAC96)

Nouns are hard to distinguish in terms of subject or object in electronic un-parsed texts such as the *CSE*. Hence, the 16.2% that are coordinated e.g. in Table 1 constitutes all functions. In Table 4, they are divided between subject and non-subject function. I have only considered cases where *faculty* is clearly the head, as in (5), and have also taken out the uses of *faculty* that were clearly modifying, as in (2) above:

(5)   One of these barriers is that *faculty* do not know what *other faculty* are doing.   (*CSE*-FAC97)

This Table shows that nouns such as *faculty* are significantly more often co-ordinated in non-subject position than in subject position, unlike (first and third person singular) pronouns where the difference between nominative and accusative is not significant.

Having shown where case or position is relevant in the *CSE* and where it is not, I now turn to the issue person. As mentioned, I will not deal with second person pronouns. The results for all first and third person singular pronouns are given in Table 5. I have added the case designations as well.

Table 5 shows that third person is more often coordinated than first person. I think the reason is that pronouns are becoming agreement markers and that the first person is 'ahead' in this respect. I'll come back to this below.

Table 5. Coordination and person in the CSE-FAC

		Uncoordinated	Coordinated	Total
first	NOM	2704	18	
	ACC	**320**	**2**	
total 1st		3024	20 (= .66%)	3044
third	NOM	180	3	
	ACC	**47**	**1**	
total 3rd		227	4 (= 1.73%)	231
total	*1+3*	3251	24 (= .73%)	3275

The figures presented in this section indicate (a) that the pronoun versus noun difference is very definite, as shown in Table 1, for all first and third person singular pronouns against four frequent nouns. The data also indicate (b) that pronominal subjects (i.e. nominatives) and non-subjects (i.e. accusatives) are almost as likely to be coordinated, but that nominal subjects and nominal non-subjects differ. Lastly, (c) there is a person split since first person is less often coordinated than third, but this is not statistically significant. I'll now turn to an account of these data.

3. Theoretical account

What do the differences in coordination tell us? I'll argue that they indicate that pronouns have less structure than nouns. I will first provide some general background on phrase structure (but see e.g. Radford 1997 for more) before showing the difference in structure between nouns and pronouns.

Within the generative tradition (e.g. Chomsky 1986), syntactic structures are built up using general rules, such as that each phrase consists of a head (X in (6)), a complement (ZP in (6)), and a specifier (YP in (6)):

(6)
```
            XP
           /  \
         YP    X'
              /  \
             X    ZP
```
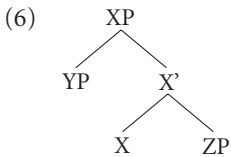
So, in (6), the X could stand for V(erb) and the YP and ZP for DPs, as in *the penguin saw a dog*. In that case, we say that a V projects up to a VP. In early work, this schema is quite strict, e.g. specifiers and complements are always full phrases, such as NPs or P(repositional) Ps, but heads are 'single words',

such as V or P. With the introduction of (minimalist) bare phrase structure in the early 1990s, this changes. A verb and a pronoun object can connect, also known as 'merge', as in (7). One of the two heads has to project, i.e. putting its characteristics on the higher phrase, in this case V:[3]

(7)
```
        VP
       /  \
      V    D
     see   it
```
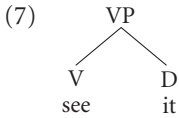
Phrase structures are built using merge and move. Merge combines two items, e.g. *see* and *it* in (7), of which one projects into a phrase. The VP domain is usually seen as the thematic-layer, i.e. where theta-roles are determined. One can think of theta-dispersion as a motivating factor behind merge, or having the structure determine them (as in Hale & Keyser 2002). After functional categories such as I(nflection) and C(omplementizer) are added to VP, a movement rule (e.g. Chomsky 1995:250) raises heads and phrases so that features (e.g. nominative case) can be checked in the IP and CP layers, as in (8), for the IP layer:

(8)
```
            IP
           /  \
          /    I'
         /    /  \
        /    I    VP
       /        /    \
      /       DP      V'
      |        |     /  \
  the penguin  V    DP
              saw  a dog
```
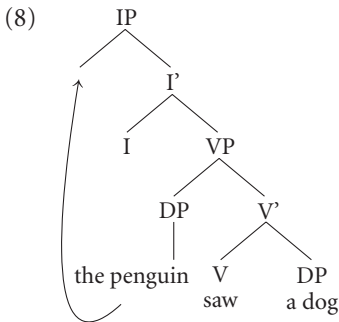
In this article, using general Minimalist principles, I argue that merging and checking between two heads is more economical than between a phrase and a head. This is formulated in van Gelderen (2004a, b) as (9), as a principle on both merge and move:

(9) *Head Preference Principle*
     'Be a Head, rather than a Phrase.'[4]

Cardinaletti and Starke (1995:36), following an older literature, analyze pronouns as being of three kinds: clitics are 'deficient heads', weak pronouns are 'deficient XPs', and strong pronouns are 'non-deficient XPs' (XPs being full phrases). In their discussion of, for instance, French, they argue that "the strong

variant can be used only if the deficient variant is not accessible" (p. 33 bold type omitted), e.g. if an adverb separates it from a verb or when coordinated. The weak pronoun "remains an XP on the surface ..., while ... resisting coordination or modification" (p. 36). Being a phrase while resisting modification seems incompatible and I will therefore reformulate Cardinaletti and Starke's three-fold distinction as a two-fold one: elements are either heads or phrases, but whenever possible they will be heads.

When coordinated or modified, pronouns are forced to be phrases. If (9) is correct, we expect them to be heads as often as possible. The way they check as heads could be through head-head checking, as in (10a), rather than spec-head checking, as in (10b), hence the frequent cliticization of the auxiliary:

(10) a.

```
                IP
          ┌─────┴─────┐
          I           VP
      ┌───┴───┐    ┌───┴───┐
      D       I   t_i       V
     she_i   's           going
```

b.

```
                IP
          ┌─────┴─────┐
         DP           I'
        she_i     ┌───┴───┐
                  I        VP
                  is   ┌───┴───┐
                      t_i       V
                              going
```
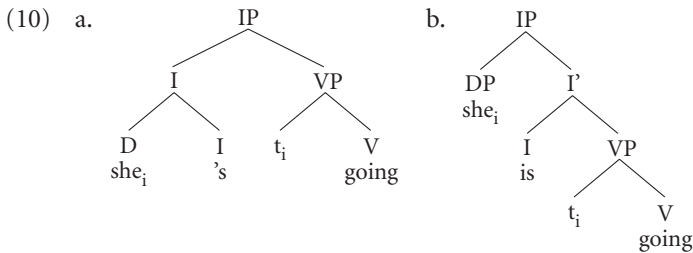
I will argue that, unlike pronouns, nouns do not have the option to be constructed as a head in (10a). Nouns, if they are to be argumental, have to have a D. Pure NPs occur as predicates, vocatives, and adverbials, but not as subjects or objects, see e.g. Higginbotham (1985), Rothstein (1983), and Longobardi (1994). This is the reason why most researchers assume a D even in languages that do not have an overt one. Comparing pronouns and nouns gives (11ab):

(11) a.

```
        D
        │
       she
```

b.

```
         DP
      ┌───┴───┐
      D       N
            coyote
```
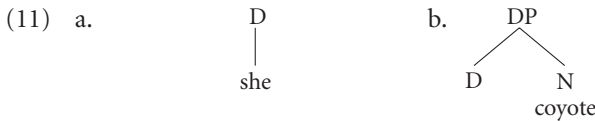
When pronouns are coordinated or modified, they lose the ability to be heads. A possible structure for a coordinate pronoun is given in (12). This is the stage in the derivation when the pronoun has just been combined with the coordinator *and*. The result is a Coordinator Phrase (as in Munn 1992; van Gelderen 1997) or a DP:

(12)

```
        DP/CoP
      ┌───┴───┐
      Co       D
     and       I
```
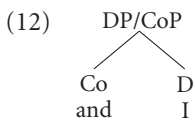
The phrase marker in (12) will still combine with a D or DP but this won't make any difference to the final result which has to be a phrase. Thus the differences between nouns and pronouns seen in Tables 1 and 2 of Section 2 can be explained through their respective structures.

I'll now turn to the theoretical relevance of the case and person data discussed in Tables 3 to 5. Table 4 shows that the subject function is less frequent with nouns than the non-subject function, whereas Table 3 shows that pronouns are more frequent in subject function. This difference is due to a discourse constraint that old or given information – and as pronouns are – typically occur in the beginning of the sentence, and new or focused information, i.e. nouns, comes towards the end. As to the coordination facts, nouns are a lot more coordinated than pronouns. However, with the pronouns there is no difference between accusative and nominative forms whereas with the nouns there is more coordination in non-subject position. This could again be the old versus new information constraint: new information is more complex.

The *CSE* never has [N(oun) *and me*] in subject position and this indicates the non-colloquial nature of the corpus. It is after all a corpus of spoken *professional* English. Prescriptive grammar says that subjects have to have nominative case, as in (13). The pied-piped preposition in (13) also exhibits this. Less formal English might have (14) with the stranded preposition. The grammatical reason is that default case is needed in coordinate phrases (see van Gelderen 1997; Johannessen 1998 for examples of default case in coordinates). Prescriptive grammar lags behind changes taking place in day-to-day speech. The data in the *CSE* exhibit that lag:

(13) the change of pace *to* which *Barbara and I* are looking forward with real relish.   (*CSE*-FAC95)

(14) the change of pace *Barbara and me* are looking forward *to*.

Sentence (13) seems awkward and this is confirmed by the figures from a much more colloquial corpus, the spoken *BNC*, that, due to its size, is harder to search. In the *BNC*, there are 304,612 instances of *I* with 656 coordinated (= .2%). There are 27703 instances of *me* with 492 coordinated (= 1.8%), so a huge difference exists between nominatives and accusatives. I will come back to this in Section 4.

Returning to Tables 3 and 4, if checking the case of subjects was different from checking that of objects, we'd expect a difference. For instance, subject pronouns if they indeed incorporate, as in (10a), would be expected to show less coordination, but this is not borne out by the data in this corpus. So,

the checking seems similar. As mentioned, non-subject nouns are more often coordinated than subject nouns, as shown in Table 4, due to discourse reasons.

The difference in person, as exemplified in Tables 1 and 5, shows that first person is more likely to be a head than third person. Van Gelderen (2000) shows that in Old English, the first person is the first to lose the pro-drop possibility and also lacks agreement on the verb. This means that the Old English first person pronoun subject is seen as a real argument whereas the other subjects are adjuncts and the 'agreement' counts as argument. If the trend seen in Table 5 is correct, this shows that again the first person is changing ahead of second and third person and is developing into agreement.

4.   Principles and change: Pronouns as agreement markers

In Section 4.1, I will look at the 1623 *First Folio* edition of Shakespeare's plays. The e-version I use (from the *Oxford Text Archive*) contains all plays and keeps the spelling and grammar of the original. This text demonstrates that around 1600 the noun/pronoun split exists, as well as a person and case one. I will not discuss the clitic situation since that depends heavily on the compositor working in the publishing house (cf. Hinman 1963). Comparing *CSE* and the Shakespeare materials shows that both are in accordance with principle (9) above. This is a good indicator that principles do not change. The slight difference is either due to the texts used (see Section 4.2), since both *CSE* and *F1* are archaic, or to a change in the status of pronouns.

4.1 Shakespeare

The figures for a few nouns as against first person pronouns are given in Table 6. These nouns are not typically used as modifiers or as verbs (even though Shakespeare is famous for converting nouns into adjectives and verbs). Comparing the figures for the nouns with those in Table 1, the difference between Shakespeare and Modern English is significant (at $p < 0.001$, χ^2 25.298). The difference in pronoun use is also significant (at $p < .001$, χ^2 12.880), i.e. pronouns are less often coordinated in the earlier text, so are used as heads more often.

The data on a possible person split are provided in Table 7, the difference not being significant between first and third person. Again a comparison with Table 5 shows that the Shakespeare data contains fewer coordinated pronouns, hence more heads.

Table 6. First person pronouns vs. nouns in F1 Shakespeare

	Uncoordinated	Coordinated	Total
I/me(e)	21291	57 (= .26%)	21348
sonne, wife, seruant, daughter	390	26 (= 6.25%)	416

$\chi^2 = 384.476$, p < 0.001.

Table 7. Pronoun and person in Shakespeare's F1 (singular and plural; nominative and accusative)

	Uncoordinated	Coordinated	Total
first	24921	66 (= .26%)	24987
third	13186	43 (= .33%)	13229

Table 8. Pronouns and case in Shakespeare's F1

		Uncoordinated	Coordinated	Total
1+3 NOM		27132	66 (= .24%)	27198
	1-NOM	18896	43 (= .23%)	18939
	3-NOM	8236	23 (= .28%)	8259
1+3 ACC		10975	43 (= .39%)	11018
	1-ACC	6025	23 (= .38%)	6048
	3-ACC	4950	20 (= .4%)	4970

There is a very noticeable case split, as shown in Table 8, for first and third person. The difference is statistically significant at p < 0.02 ($\chi^2 = 6.007$). As we'll see later, this brings the figures in the same range as those of the *BNC*, and confirms that accusative pronouns have more often been coordinated, at least since 1600. If accusative is the default case and if pronouns in coordinate phrases get default case, this fits since they don't check case in a configuration where being a head is more economical.

So far, it has been shown that Shakespeare's plays show the same tendencies as the Modern English *CSE*. This is accounted for by (9).

The reason that (9) is not predicting a change is that it is a principle. Of course, there is the possibility that the differences noticed may have to do with the texts used. As mentioned, in the case of the *CSE*, even though it comprises transcripts of spoken American English, the English is very formal with the result that very few 'errors' occur. For instance, *me* is never used in a subject coordinate, whereas in the *BNC* that use, as in (15), is a lot more frequent than

Table 9. First person singular pronouns in HC, F1, BNC, and CSE

	Coordinated/total nominative		Coordinated/total accusative		Coordinated/total all	
HC	45/8166	(= 0.55%)	34/1446	(= 2.35%)	79/9612	(= 0.82%)
F1	43/18939	(= 0.23%)	23/6048	(= 0.38%)	66/24987	(= 0.26%)
BNC	656/304612	(= 0.2%)	492/27703	(= 1.8%)	1148/332315	(= 0.34%)
CSE	18/2722	(= 0.66%)	2/322	(= 0.62%)	20/3044	(= 0.66%)

the nominative *I* (interestingly [*me and* N] is more frequent than [N *and me*], going against another prescriptive rule and the same holds for the nominative):

(15)   *me and my mother* have erm arranged it all   (*BNC*-KC8 920)

The *BNC* is much less formal than the *CSE* and will therefore be used next. Shakespeare's plays may also not be the right kind of text and I have therefore examined the Early Modern English section (EMOD1-3) of the *Helsinki Corpus,* which contains a more balanced set of texts.

4.2 *BNC* and *HC*

Putting the figures from the *BNC* and the *Helsinki Corpus* (EMOD1-3) section together with the earlier *CSE* and the Shakespeare figures shows the following for first person nominative and accusative. Later research will have to consider third person.

Comparing the *HC* (made up of formal and informal writings) and the spoken *BNC* reveals that there is a change towards less coordination where pronouns are concerned. The difference between the *BNC* and *HC* is statistically significant (at $p < 0.001$, $\chi^2 = 39.776$). The reason that the *CSE* and the *F1* text are contrasting has probably to do with the kind of English used. It could be argued that the English of the *CSE* is artificial, especially where pronouns are concerned. This brings up the more general question of the nature of data, as addressed in Weiss (2001) and Schütze (1996). Weiss argues that standard languages, learned by special instruction, e.g. in schools, have properties which may not entirely be due to Universal Grammar. The language used in the *CSE* may be an example of such a language.

In conclusion to Section 4, the contrast between nouns and pronouns is robust in all texts of all periods. This is explained by (9) being a principle in speakers' grammars. Depending on the naturalness of the corpus, the use of pronouns as heads can also be seen to have increased over time. Nominative pronouns are less often coordinated in texts of all periods, presumably because

the accusative is a default case, used if case cannot be assigned properly (cf. van Gelderen 1997; Johannessen 1998). The difference in person is not investigated in all corpora but in the *CSE*, the first person is most likely a head, possibly becoming agreement (see Table 5) and in the *BNC*, *he* is coordinated in 4% of the cases whereas *I* in only 0.2% (these percentages are based on samples since the numbers are so large).[5] As mentioned, it is interesting in this respect that first person loses agreement on the verb early on and is least likely dropped in pro-drop in Old English. With the contemporary change from pronoun to agreement marker, the first person is again the frontrunner. The reason for this change is not clear but shows that it is not a functional principle, as e.g. in Ariel (2000) who argues that first person has pro-drop more frequently because of the accessibility of the referent.

5.   Parametric change is fast

The changes I will now discuss are not related to the Head Preference Principle, but to the loss of morphological case, a parameter setting change from morphological case to checking case in a Functional Category such as I(nflection). This change is fast, or 'catastrophic', as Lightfoot (1999) calls it. I show that first person is the first to lose the special morphological case. This has to do with checking and reveals another change going on in Old English 'favouring' first person. Splits in corpora data cannot tell us anything about the change since it is very fast.

In Old English, nouns and pronouns can be marked genitive, dative, or accusative depending on the character of the governing verb. This case is not related to position but to theta-marking (cf. Kiparsky 1995). Chomsky (1995) refers to it as inherent case, as opposed to structural case which is related to positions such as the specifier of the IP. By 1200, the different cases in English fall together (at least morphologically), and just nominative and accusative are used (for sentential arguments). As a result of the change to structural case, word order becomes stricter since e.g. subjects have to be in certain structural positions. This change in case is related to the changes discussed in Section 2. If agreement marking in Old English is actually an argument, as in pronominal argument languages, checking in a higher functional category is not necessary. As endings (both agreement and case) disappear, checking becomes necessary and overt arguments occur. I now show what corpora say about the case change.

Table 10. First person changes in the Helsinki Corpus

	OE1-2 (–950)	OE3 (950–1050)	OE4 (1050–1150)	ME1 (1150–>)
me	597	1282	234	669
mec	90 (= 13%)	194 (= 13%)	0	0
total	687	1476	234	669

As shown in Table 10, between OE1-2 and OE3 of the *HC*,[6] i.e. around 1000, the relationship between the specially marked accusative *mec* and the dative/accusative *me* remains stable. An example with both is given in (16) from Mercian *Vespasian Psalter*:

(16)   ða   ðe   swencað _mec_...   monge arisað wið _me_
      that that oppress me-ACC many rise   with me-DAT
      "that oppress me ... many rise with me".

<div align="right">(Vespasian Psalter, 3.1, Kuhn edition)</div>

By OE4, again, shown in Table 10, i.e. from 1050 on, *mec* has disappeared, however. So, even though the ratio stays the same between OE1 and OE3, one form suddenly disappears. This means that the change is rapid and the corpus data does not help predict it.

There may be problems in using the OE part of the *HC* (see Note 6). Using the complete Old English corpus (*DOE*) is, however, even harder. The percentage of *mec* compared to all accusative and dative forms is 6.3%. Even though in Modern English (e.g. in the *BNC*), the percentage of *me* as opposed to *I* is much lower, we would not predict the demise of *me* and the survival of *I*. Hence, one could not have predicted the instability of *mec* either.

Third person data are more difficult to work with since the forms are so numerous. For instance, *he* and *him* are spelled in many different ways and the latter represents different numbers. The accusative *hine/hyne* is only used for masculine singular and that makes the comparison harder. In Tables 11a/b, I have compared the general 'dative' with the accusative (and have ignored e.g. *heom*). The latter's demise seems to occur between ME1 and ME2.[7]

So, just based on the variety of the forms, Tables 10 and 11a/b demonstrate that morphological case is disappearing. I'll now show that a special (impersonal) construction, making use of morphological case, exhibits the same direction of change.

Apart from morphology, another indicator that shows that morphological case is disappearing is the loss of constructions such as (17) and their replacement by (18), a slightly more modern version of the same original:

Table 11a. Third person changes in the HC (OE1-2–>OE4)

	OE1-2 (–950)	OE3 (950–1050)	OE4 (1050–1150)
him	922	2166	631
hine	411 (= 31%)	985 (= 31%)	259 (= 29%)
total	1333	3151	890

Table 11b. Third person changes in the HC (ME1–>ME3)

	ME1 (1150–1250)	ME2 (1250–1350)	ME3 (1350–1420)
him/hym	1054	1158	1544
hine/hyne	209 (= 17%)	10 (= .86%)	0
total	1263	1186	1544

(17) *þer-fore* <u>*him*</u> *ofte   scomede. 7   his heorte gromede*
     therefore him often shamed and his heart angered
     "therefore he often felt ashamed and enraged".   (Caligula 6868)

(18) *þar-fore* <u>*he*</u> *ofte   samede. and hi   heorte gramede*
     therefore he often shamed and his heart angered
     "therefore he often felt ashamed and enraged".   (Otho 6868)

From Jespersen (1894) and van der Gaaf (1904) on, (17) has been referred to as an impersonal. It could also be seen as ergative, since the 'subject' has object characteristics.

As expected, considering the changes displayed in Tables 10 and 11a/b, the first person is the frontrunner in the change from (17) to (18) as well. The ratio of impersonal or ergative uses out of the total number of dative forms for all persons together results in Table 12 for *Beowulf* (taken from van Gelderen 2000: 238), an early Old English text. Statistically, the difference between 1S and 3S (χ^2 is 5.342, $p < 0.005$), and between 1S/2S versus 3S (χ^2 12.018, $p < 0.001$) is significant, but not between 1S and 2S, which is not unexpected. Only for third person is number statistically relevant.

An alternative to Table 12 is Table 13 (again taken from van Gelderen 2000: 238), with the percentage of impersonal or ergative use as against the personal or non-ergative use of the nominative form. The difference between 1S and 3S (χ^2 is 14.269, $p < 0.001$), and between 1S/2S versus 3S (χ^2 17.285, $p < 0.001$) is significant, but, as before, not between 1S and 2S. Again only for third person is number statistically relevant.

Table 12. Ergative versus non-ergative uses of first, second and third person dative pronouns in 'Beowulf' (S = singular; P = plural)

	Ergative	Non-ergative	Total
first-S	7 (= 12.7%)	48	55
second-S	3 (= 6.8%)	41	44
third-S	47 (= 28.1%)	120	167
first-P	1 (= 10%)	9	10
second-P	1 (= 16.7%)	5	6
third-P	3 (= 9.1%)	30	33
total	62 (= 19.7%)	253	315

Table 13. Ergative versus nominative in 'Beowulf'

	Ergative	Nominative	Total
first-S	7 (= 3.7%)	181	188
second-S	3 (= 4.6%)	62	65
third-S	47 (= 14.3%)	282	329
first-P	1 (= 4%)	24	25
second-P	1 (= 7.7%)	12	13
third-P	3 (= 4%)	71	74
total	62 (= 13.9%)	632	694

The problem with this Table is that third person pro-drop is more prevalent than first and second person and that would reduce the percentage for third person anyway (cf. e.g. Berndt 1956 and van Gelderen 2000).

To conclude Section 5, I have briefly shown that the specialized case system is lost, starting with the first person. This change is fast and due to a parameter resetting, and not predictable by shifts in the corpora. I would also argue that this change is, in fact, connected to that discussed in Section 4. Since first person is the first to lose verbal agreement and pro-drop, i.e. pronominal argument-hood, it is the first to start checking in a functional category such as I (see (10) above). The modern change towards incorporating the first person in the verb is just another step in that cycle.

6.   Conclusion

In this paper, I have shown how corpora can be used to 'get at' phenomena not (yet) very visible. Pronouns and nouns in Modern English behave differently where coordination is concerned and the reason for this is the Head

Preference Economy Principle. Nominative and accusative pronouns also behave slightly differently, as do the different persons, at least in the *BNC*. The differences regarding the pronoun-noun split between the *HC* and the *BNC* are very slight and I argue this is due to the Head Preference Economy Principle being a principle rather than a parameter. The paper demonstrates that changes due to the resetting of parameters are different in nature from those due to principles: principles shouldn't change but parameters can be reset. With the latter change, they are characteristically fast, 'catastrophic' in Lightfoot's terms, and not predictable in a corpus. The parameter setting examined is the loss of morphological case. The loss of this case means that case becomes checked in inflectional categories (hence inducing stricter word order in Modern English).

Notes

* Thanks to Chen Chen Sun, Leonie Cornips, Karen Corrigan, Johanna Wood, and two anonymous referees for helpful comments.

1. There are a number of aspects I do not go into. For instance, pronouns can be emphatic or in focus, i.e. the italicized pronouns in (i) and (ii) respectively. I assume these to be full phrases, but I have not looked at them:

(i)   *Me,* I don't want to do that.

(ii)   *She* perhaps might be doing that.

2. The numbers for coordinated nouns and pronouns include 'and X' as well as 'X and'. I have not taken into account instances where 'and' functions as an adverb introducing new sentences. For Table 1, I have selected the nouns in the text that are the most frequent, and have not taken into consideration their status as mass or count noun.

3. Strictly speaking, bare phrase structures have no labels, but I will continue to use them for convenience.

4. It may be that just moving features is even more economical, as Chomsky (1995:262ff.) suggests, or indeed involving 'agree' between a probe and goal, as in later work. Stated as in (9), the principle holds for merge (projection) as well as move (checking).

5. I examined the first 1000 instances of each.

6. As a reviewer points out, *Beowulf* is a poetic text and so are some of the texts in the HC; others are glosses. Some people feel these don't represent Old English very well. There are some problems in using the HC the way I have since the OE periods are based on the dates of the manuscripts, not necessarily on the dates of composition.

7. Plural pronouns are less often coordinated for reasons I don't go into. For instance, first person plurals occur 2561 times, of which 3 (or .12%) are coordinated. Third person ones occur 658 times and are coordinated only once (or .15%).

References

Ariel, Mira (2000). "The Development of Person Agreement Markers". In Michael Barlow et al. (Eds.), *Usage-based Models of Language* (pp. 197–260). Stanford: CSLI.

Bergen, Linda van (2003). *Pronouns and Word Order in Old English*. Abingdon, Oxford: Routledge (Taylor and Francis).

Berndt, Rolf (1956). *Form und Funktion des Verbums im nördlichen Spätaltenglischen*. Halle: Max Niemeyer.

British National Corpus (*BNC*). Available online at <thetis.bl.uk.> [100 million words of British English, 10% spoken].

Brook, G. L. & R. F. Leslie (Eds.). (1963). *Layamon: Brut*. Oxford: Oxford University Press, EETS 250. [Two Manuscripts: *Caligula* and *Otho*, now both said to be from the second half of the 13th Century, (N) Worcestershire].

Campbell, Alistair ([1959] 1997). *Old English Grammar*. Oxford: Clarendon Press.

Cardinaletti, Anna & Michael Starke (1996). "Deficient Pronouns: A View from Germanic". In Höskuldur Thráinsson et al. (Eds.), *Studies in Comparative Syntax II* (pp. 21–65). Dordrecht: Kluwer.

Chomsky, Noam (1962). "Discussion". In Archibald Hill (Ed.), *Third Texas Conference on Problems of Linguistic Analysis in English* (pp. 124–169). Austin: University of Texas.

Chomsky, Noam (1986). *Barriers*. Cambridge: MIT Press.

Chomsky, Noam (1995). *The Minimalist Program*. Cambridge: MIT Press.

Cornips, Leonie & Karen Corrigan (2005). "Convergence and Divergence in Grammar". In Peter Auer et al. (Eds.), *Dialect Change: Convergence and Divergence in European Languages* (pp. 96–134). Cambridge: Cambridge University Press.

Corpus of Spoken Professional American English (*CSE*). Not available online, but see <www.athel.com.> [2 million words; divided in COMM, FAC, WH].

Gaaf, W. van der (1904). *The Transition from the Impersonal to the Personal Construction in Middle English*. Heidelberg: Carl Winter.

Gelderen, Elly van (1997). *Verbal Agreement*. Tübingen: Niemeyer.

Gelderen, Elly van (2000). *A History of English Reflexive Pronouns*. Amsterdam: John Benjamins.

Gelderen, Elly van (2004a). "Specifiers, Heads, Grammaticalization, and Economy". *Journal of Comparative Germanic Linguistics, 7*(1), 59–98.

Gelderen, Elly van (2004b). *Grammaticalization as Economy*. Amsterdam: John Benjamins.

Hale, Ken & Jay Keyser (2002). *Prolegomenon to a Theory of Argument Structure*. Cambridge: MIT Press.

Helsinki Corpus (*HC*). [Available through the *Oxford Text Archive*.]

Henry, Alison (1995). *Belfast English and Standard English*. Oxford: Oxford University Press.

Higginbotham, James (1985). "On Semantics". *Linguistic Inquiry, 16*, 547–593.

Hinman, Charlton (1963). *The Printing and Proofreading of the First Folio of Shakespeare*. Oxford: Oxford University Press.

Jespersen, Otto (1894). *Progress in Language*. London: Allen & Unwin.

Johannessen, Janne (1998). *Coordination*. Oxford: Oxford University Press.

Kayne, Richard (1989). "Facets of Romance Past Participle Agreement". In Paola Benincà (Ed.), *Dialect Variation and the Theory of Grammar* (pp. 85–103). Dordrecht: Foris.

Kiparsky, Paul (1995). "Indo-European Origins of Germanic Syntax". In Adrian Battye & Ian Roberts (Eds.), *Clause Structure and Language Change* (pp. 140–169). Oxford: Oxford University Press.

Kroch, Anthony & Ann Taylor (Eds.). (2000). *Penn-Helsinki Parsed Corpus of Middle English* (2nd edition). Philadelphia: Department of Linguistics, University of Pennsylvania. [Standard abbreviation: PPCME2.]

Kuhn, Sherman M. (Ed.). (1965). *The Vespasian Psalter*. Ann Arbor: University of Michigan Press.

Lightfoot, David (1999). *The Development of Language*. Malden, MA: Blackwell.

Longobardi, Giuseppe (1994). "Reference and Proper Names". *Linguistic Inquiry, 25*, 609–665.

Munn, Alan (1992). "A Null Operator Analysis of Parasitic Gaps". *The Linguistic Review, 9*(1), 1–26.

Pintzuk, Susan (1999). *Phrase Structures in Competition*. New York: Garland.

Pintzuk, Susan, Eric Haeberli, Ans van Kemenade, Willem Koopman, & Frank Beths (Eds.). (2000). *The Brooklyn-Geneva-Amsterdam-Helsinki Parsed Corpus of Old English*. York: Department of Language and Linguistic Science, University of York. [Standard abbreviation: The Brooklyn Corpus.]

Radford, Andrew (1997). *Syntactic Theory and the Study of Syntax*. Cambridge: Cambridge University Press.

Rothstein, Susan (1983). The Syntactic Forms of Predication. Ph.D. dissertation, MIT.

Schütze, Carson (1996). *The Empirical Base of Linguistics*. Chicago: University of Chicago Press.

Shakespeare, William (1623). *First Folio* in electronic form. [Available through the *Oxford Text Archive.*]

Taylor, Ann, Anthony Warner, Susan Pintzuk, & Frank Beths (Eds.). (2003). *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. York: Department of Language and Linguistic Science, University of York. [Available through the *Oxford Text Archive.*]

Trips, Carola (2002). *From OV to VO*. Amsterdam: John Benjamins.

Wasow, Tom (2002). *Postverbal Behavior*. Stanford: CSLI.

Weiss, Helmut (2001). "On Two Types of Natural Languages". *Theoretical Linguistics, 27*, 87–103.

Wilson, J. & Alison Henry (1998). "Parameter Setting within a Socially Realistic Linguistics". *Language in Society, 27*, 1–21.

Wood, Johanna (2003). Definiteness and Number. Ph.D. dissertation, ASU.

Morphosyntactic variation and theory
Subject-verb agreement in Acadian French*

Ruth King
York University

1.   Introduction

This paper contributes to the enterprise of integrating formal and sociolinguistic perspectives by providing an analysis of a particular case of morphosyntactic variation which combines the two. On the one hand, this work is part of the sociolinguistic tradition of finely-grained quantitative analysis of grammatical variation found in large sociolinguistic corpora; on the other, it is couched within the generative tradition in that it seeks to elucidate microvariation by analysing very closely-related grammatical systems using the technical apparatus that framework makes available. I first contextualise this approach in terms of the treatment of grammatical variation in variationist linguistics and in terms of the small but growing body of research which combines the two perspectives. I then turn to a brief description of Acadian French, (marginalised) French varieties spoken in Atlantic Canada, and the corpora on which the present study is based. In the bulk of the paper I provide an account of third person plural marking in Acadian varieties, in particular the Newfoundland variety, with the aim of showing that combining the insights of generative grammar with traditional sociolinguistic analysis provides a more complete picture than analysis along one dimension alone would offer.

2.   Morphosyntactic variation and change

In a chapter of his landmark work *Language in the Inner City*, William Labov (1972) analyzes negative concord in African American Vernacular English,

countering the naive view that a striking contrast exists between Standard English and AAVE, the latter thought to have "too many negatives." Labov shows instead that the distribution and interpretation of negatives illustrated by the often-cited example in (1) – which would be interpreted as (2) by speakers of Standard English but as (3) by AAVE speakers – is not a case of an unconstrained proliferation of negatives.[1]

(1)   It ain't no cat can't get in no chicken coop

(2)   There isn't any cat that cannot get into any chicken coop

(3)   There isn't any cat that can get into any chicken coop

Rather, close inspection reveals that negative concord in the two varieties is not really very different; instead, variation "may be traced to minimal adjustments in English phonological and grammatical rules" (Labov 1972:131). I would argue that Labov's early work on grammatical variation, while it might differ in the technicalities, is actually in the spirit of current analyses of variation which combine formal and sociolinguistic perspectives, including my own.

While research on grammatical variation within the Labovian tradition dates back well over thirty years, there is a tendency for outsiders and some insiders as well to consider variationist linguistics to be much more centrally focussed on the study of phonological variation and change. However, as Labov (1994) has argued convincingly, this is in fact not so: rather, studies of grammatical variation have actually been more numerous than studies involving phonetics/phonology. As evidence for this claim Labov calculated the proportion of papers devoted to the various subdisciplines of linguistics presented at several NWAVE conferences in the 1970s and 1980s and also the proportion of papers published in the journal *Language Variation and Change*, the major outlet for variationist linguistics in North America. For both sources, studies of grammatical variation were in fact more frequent.[2]

While it is the case that, as Henry (2002:277) has suggested, some of this work is uninformed by any particular theory of grammar, more functional approaches to language have been prominent in studies of grammatical variation, such as functionalist approaches to grammaticalization (e.g. Poplack & Tagliamonte 2001; Sankoff 1990) and to the relationship between syntax and discourse/pragmatics (see Sankoff 1988 for discussion). The study of diachronic syntax, however, has seen the rise of a particularly rich tradition of combining quantitative and formal perspectives, dating from the 1980s (see Kroch 2000 for an overview). There is also a small but important body of research which, over the past decade or so, has attempted to combine the

insights of generative linguistics with variationist analysis in the study of synchronic variation (cf. Auger 1998; Cornips & Corrigan 2005; Heap 2001; King & Nadasdi 1997; King 2000; Meechan & Foley 1994; Meyerhoff 2000; Nadasdi 1995; Wilson & Henry 1998).

I provide below an example of this fruitful combination of variationist and generative approaches by focussing on variation in third person plural marking in a number of related varieties of French, showing that the analysis of this case of variable agreement, while partially couched in recent generative theory, is not far removed from Labov's early research. Further, I show that a full account of variable number marking benefits strongly from a turn to formal theory, in this case the Minimalist Program (Chomsky 1993, 2001) along with the theory of Distributed Morphology (Halle & Marantz 1993).

3.   Acadian French

Acadian French refers to (marginalised) varieties of French spoken principally in the Canadian provinces of New Brunswick, Newfoundland, Nova Scotia, and Prince Edward Island (see Map 1).[3]

Acadian differs from its more well-known neighbour, Quebec French, due (in part) to the different European origins of the colonists, with the majority of Acadian settlers coming from the provinces of the *centre-ouest* of France (Map 2), whereas Quebec colonists were more diversified, with substantial numbers of settlers from north of the Loire Valley.

French settlement in North America dates from the early years of the 17th century, with the beginnings of Acadia dating from 1605, when the French explorer Samuel de Champlain and the French nobleman Pierre du Gast, Sieur de Monts, established a colony on the shores of the Baie Française (the Bay of Fundy). By the end of that century, the population of Acadia (the place-name was originally applied to peninsular Nova Scotia) consisted mainly of second-to-fourth generation settlers, who formed a social group cohesive enough to allow historians to refer to them from this point on as the Acadian people. However, by the Treaty of Utrecht in 1713, England was given control over the area, at a time when the local population numbered between 1500 and 2000 while France retained what is now Prince Edward Island, Cape Breton Island (part of the present-day province of Nova Scotia), and the coastline of present-day New Brunswick. The French government invited former colonists to settle these areas and a large number did so, leaving British jurisdiction. De-
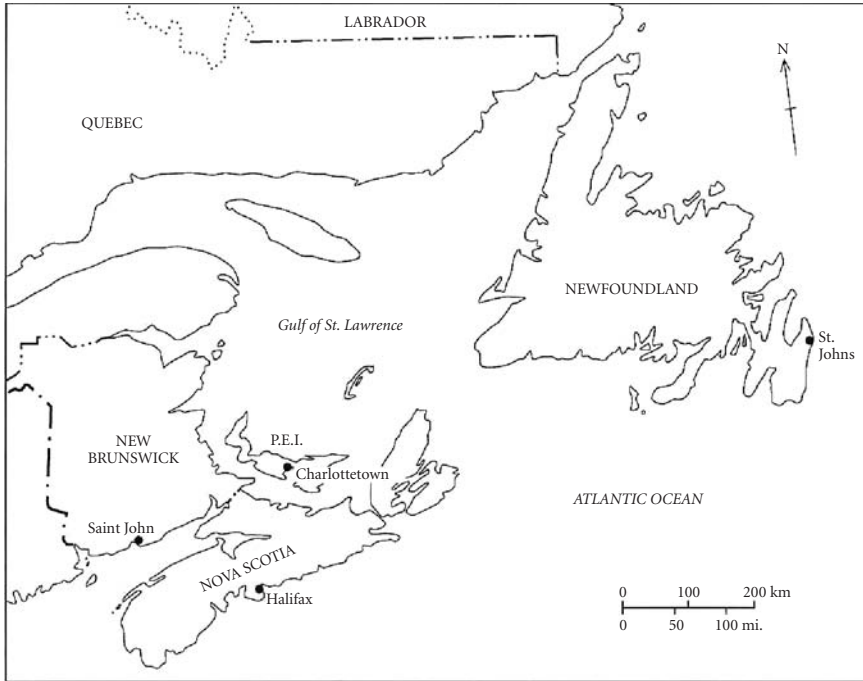
Map 1. Canada's Atlantic Provinces (with provincial capitals indicated)

spite the exodus, the French population of the Acadian peninsula would reach approximately 14,000 by the outbreak of the Seven Years' War in 1756.

In 1755, the colonial governor of Nova Scotia ordered the deportation of the Acadians from the British colony, considering them a security risk given the strategic location of the Acadian peninsula in time of war. Between 1755 and 1765, the British deported the Acadians far and wide, to the New England colonies, to Louisiana, then a French colony governed by Spain, to the West Indies, to France, etc., with their lands taken over by Anglo-American settlers from New England. Of those who escaped the Deportation, many Acadians fled into the wilderness of what is now New Brunswick and to Prince Edward Island and to the Gaspé Peninsula, part of present-day Quebec. Following the fall to the British of Louisburg on Cape Breton Island to the British in 1758, the Acadian settlements there on Cape Breton Island and on Prince Edward Island, too, were destroyed.

While the years following the Deportation resulted in the dispersal of the Acadian people, the return from exile began in the late 1760s and lasted a quar-

Map 2. The *centre-ouest* provinces of France

ter of a century, with groups of Acadians settling in isolated areas of Nova Scotia and Prince Edward Island. They could not reclaim their original, fertile lands but were left with what was essentially the land that the English did not want. The years following saw Acadians establishing isolated settlements where land was available, often surrounded by English settlements. This relative isolation of Acadians from contact with other Francophones and from the normative influences of a French education system (this was now a British colony) is perhaps the most important reason for differences between Acadian and Quebec French, with Acadian retaining many conservative features of the language. This isolation lasted into the 20th century for some varieties and continues to the present day for others (Arsenault 1987; Butler 1995; Flikeid 1994; Ross & Deveau 1992). The present-day situation ranges from one in which French is the majority language in north eastern New Brunswick, with Francophones comprising a third of the province's population as a whole, to the existence of scattered Acadian settlements in the other Atlantic Provinces, with varying degrees of language maintenance.

Sociolinguistic studies of present-day Acadian varieties reveal complex, but structured, organization of linguistic variation, or as Labov has termed it, orderly heterogeneity. Several studies reveal a tension between maintenance of Acadian linguistic features (markers of Acadian identity) and linguistic change in the direction of community-external standards. For example, a number of researchers (e.g. Flikeid & Péronnet 1989; King, Nadasdi, & Butler 2004) have studied first person plural pronominal variation, documenting the rise of the general French colloquial variant *on* at the expense of the traditional first person form *je*, unmarked for number (e.g. *je parle* "I speak"; *je parlons* "we speak"), in Acadian communities with substantial contact with external varieties of French. Another source of variation involves length and degree of contact with English, the majority language in most regions of Atlantic Canada, which accounts for intercommunity differences in the linguistic consequences of such contact (e.g. Flikeid 1989; King & Nadasdi 1999). For example, while all Acadian varieties have borrowed verbs from English (and morphologically incorporated them into the recipient language), only those varieties in most intense content with English have also borrowed functional elements such as prepositions. Further, the borrowing of functional elements can be seen to have consequences for the grammars of contact varieties (King 2000).

While Acadian French is the site of linguistic innovation, it has also proven to provide a window on the past. For example, in the Newfoundland and Prince Edward Island varieties, subject pronouns remain syntactic subjects, as opposed to prefixes on the verb (King & Nadasdi 1997): i.e. they have not undergone a change from clitic to affixal status largely agreed to have occurred in most varieties of colloquial French (cf. Auger 1994). However, such a change has arguably taken place in one New Brunswick variety, which shows innovation along the lines of Quebec French (Beaulieu & Balcom 1998). Such cases of striking interdialectal variation point to the need to proceed with in-depth analysis one variety at a time in order to avoid premature conclusions about just what constitutes "Acadian French".

The study of Acadian French, then, involves the study of closely-related language varieties which have similar grammars but which vary in terms of their social circumstances. As such, it provides an ideal testing ground for theories of the social mechanisms of linguistic change. And since these grammars differ minimally from each other, they readily allow for the study of microvariation. In what follows, I combine variationist and generativist approaches in the examination of a particular morphosyntactic phenomenon, third person plural agreement marking, in three Acadian varieties.

While the data for the analysis come from sociolinguistic interview cor-
pora for several Acadian communities, the more detailed linguistic analysis
draws mainly on data for an Acadian variety spoken in Newfoundland (Map 3).
French settlement came fairly late to Newfoundland, dating from the 19th cen-
tury immigration of Acadians from the Chéticamp area of Nova Scotia (Map
4) and of a small number of metropolitan French (Butler 1991; King & Butler
in press). In general, the Newfoundland varieties are among the most conser-
vative of Acadian varieties, having little contact with English until World War
II, and, even after that, they remained comparatively isolated until the 1970s
(King & Butler in press).[4] The Newfoundland corpus used here is for the vil-
lage of L'Anse-à-Canards, a traditional fishing village of less than 200 people;
the village itself is located on the province's isolated Port-au-Port peninsula,
as noted above the only area of the province with Acadian settlements. Most
importantly for the present study, there was – and still is – relatively little con-
tact with external varieties of French. The Newfoundland corpus consists of
approximately one million words, of sociolinguistic interviews, both one-on-
one interviews and group interviews, conducted by native speaker residents in
the late 1980s. For comparative purposes, I have examined usage in a slightly
smaller (640,000 words for the main interview corpora) Prince Edward Is-
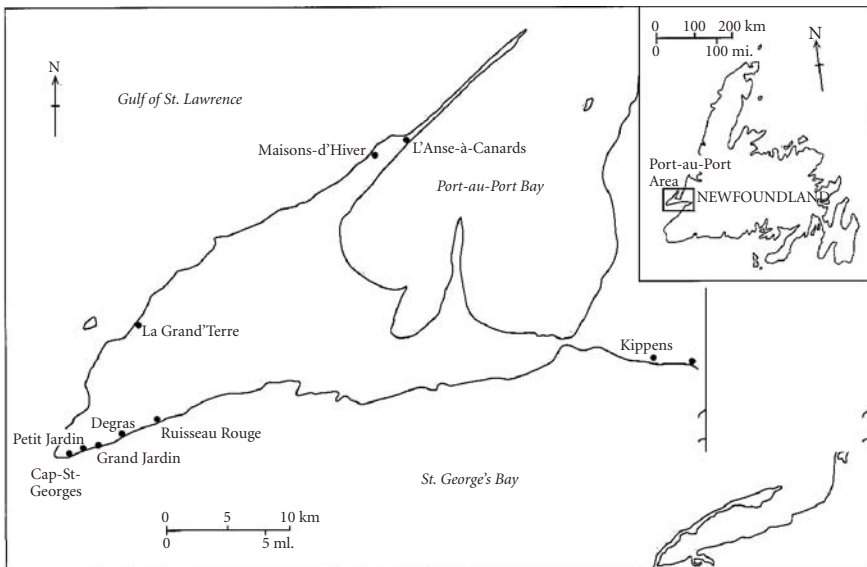
Map 3. Newfoundland Acadian Communities
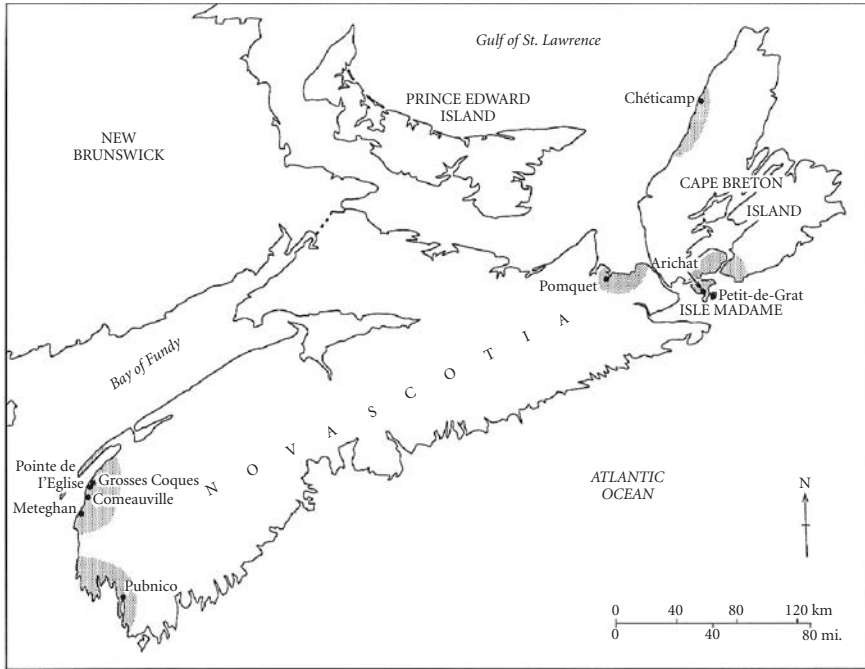
Map 4. Nova Scotia Acadian Communities

land corpus which consists of interview data from two villages, Abram-Village and Saint-Louis. Like L'Anse-à-Canards, Abram-Village was a majority French village at the time of corpus construction, also in the late 1980s, but one in which there was considerable institutional support for Standard French and exposure to external varieties. However, the second Prince Edward Island community, Saint-Louis, was one with much less contact with external varieties and with French in a minority, indeed crisis, situation within the community itself, with younger speakers ceasing to acquire the language.[5] Thus there is variation across the three corpora in terms of degree of external influence, both from other French varieties and from English.

4. Third person plural marking

Perhaps the most well-known feature of (conservative) Acadian French involves the verb morphology and the pronominal system. By way of example,

Table 1. Conjugation of the verb parler "to speak" in Acadian French

Person/Number	Present	Imperfect	Perfect
1 singular	je parle	je parlais	j'ai/as parlé
2 singular	tu parles	tu parlais	t'as parlé
3 singular	il/elle parle	il/elle parlait	il/elle a parlé
1 plural	je parlons	je parlions	j'avons/ons parlé
2 plural	vous parlez	vous parliez	vouz avez parlé
3 plural	ils parlont	ils parliont	ils avont/ont parlé

Table 1 shows the verb *parler* "to speak" as it is conjugated in conservative Acadian varieties.

The singular verb forms tend to be homophonous, as is the case in any number of colloquial French varieties, and, for regular verbs, in Standard French as well. As noted above, the first person present plural (*je parlons* "we are speaking") has an [õ] suffix. So, too, does the third person present plural (*ils parlont* "they are speaking"). The corresponding imperfect (*je parlions* "we were speaking"; *ils parliont* "they were speaking") and perfect (*j'avons/ons parlé* "we spoke"; *ils avont/ont parlé*) likewise take the [õ] suffix. Both first and third person plural forms are widely regarded as "typical" Acadian French.[6] According to Brunot (1967.2.335), the archaic *je...ons* began to disappear in French in the 16th century. While fairly infrequent in earlier Quebec French and absent from the present-day variety (King, Martineau, & Mougeon 2004), its geographical distribution remained widespread in France up until the late 19th century, as evidenced by the *Atlas Linguistique de la France,* which shows it occurring in most northern varieties. Today, however, *je...ons* is restricted to the most conservative Atlantic Canada Acadian varieties.

Third person plural *-ont* is likewise an example of archaic usage, appearing in France as early as the thirteenth century. According to Nyrop (1979.6), it was in widespread usage in the *centre-oeust* of France at the time of Acadian emigration to the New World, in the 17th century. Although much more geographically restricted than *je...ons*, its use extended north as far as Brittany and Normandy.[7] Today, in non-Acadian varieties of French, the third person singular and plural most often are homophonous. Thus in Quebec French and in European French, along with Standard French, the third person singular and plural are alike except when the verb is in the inflected future (e.g. *il aura* "he will have" – *ils auront* "they will have") or in that small number of cases in which there is alternation in the shape of the stem (e.g. *il peut* "he can" – *ils peuvent* "they can").[8] Orthographic *-ent*, then, is phonologically null except in liaison contexts. In contrast, in the traditional Acadian system, third person

singular and plural are always distinct (e.g. *il parle* "he is speaking"; *ils parlont* "they are speaking"). (4)–(7) below give examples of Acadian *-ont* taken from the L'Anse-à-Canards corpus (each example is followed by the speaker number). The third person singular forms common to all varieties and the corresponding Standard French plurals are also indicated.

(4)   Les gens   de   delà, ils   *parlont* curieux   (AC-5)
      the people from there they speak   funny
      "The people from there, they speak funny."
      (3rd sg: il parle; SF 3rd pl: ils parlent)

(5)   Ils   *voyiont*   une lumière à tous les soirs   (AC-2)
      they were-seeing a   light   at all   the evenings
      "They used to see a light every evening."
      (3rd sg: il voyait; SF 3rd pl: ils voyaient)

(6)   Ils   *savont* quoi ce-que ça   veut   dire[9]   (AC-2)
      they know   what that   that wants to-say
      "They know what that means."
      (3rd sg: il sait; SF 3rd pl: ils savent)

(7)   Ils   *avont* dit   que le   diable a   venu[10]   (AC-4)
      they have   said that the devil   has come
      "They said that the devil came."
      (3rd sg: il a dit; SF 3rd pl: ils ont dit)

5.   Degree of retention of the conservative system

The degree of retention of the traditional third person plural variant shown in (4)–(7) has been the subject of a number of prior studies. I begin by summarizing the results of the earliest of these studies, a comparison of usage in several locations in Nova Scotia (Map 4) and New Brunswick (Map 5).

   A 1989 study by Karin Flikeid and Louise Péronnet investigated third person plural subject-verb agreement for a number of Nova Scotia and New Brunswick varieties of Acadian French. Their focus was on the proportion of use of the Acadian agreement pattern in third person plurals by older speakers.[11] The results range from 70% to 87% for six locations, indicating, in general, that use of the traditional variant is robust, at least for older speakers. Flikeid and Péronnet found that while these intercommunity differences were not statistically significant, what was crucial in explaining variation was the level of education attained by the individuals involved: those with less than
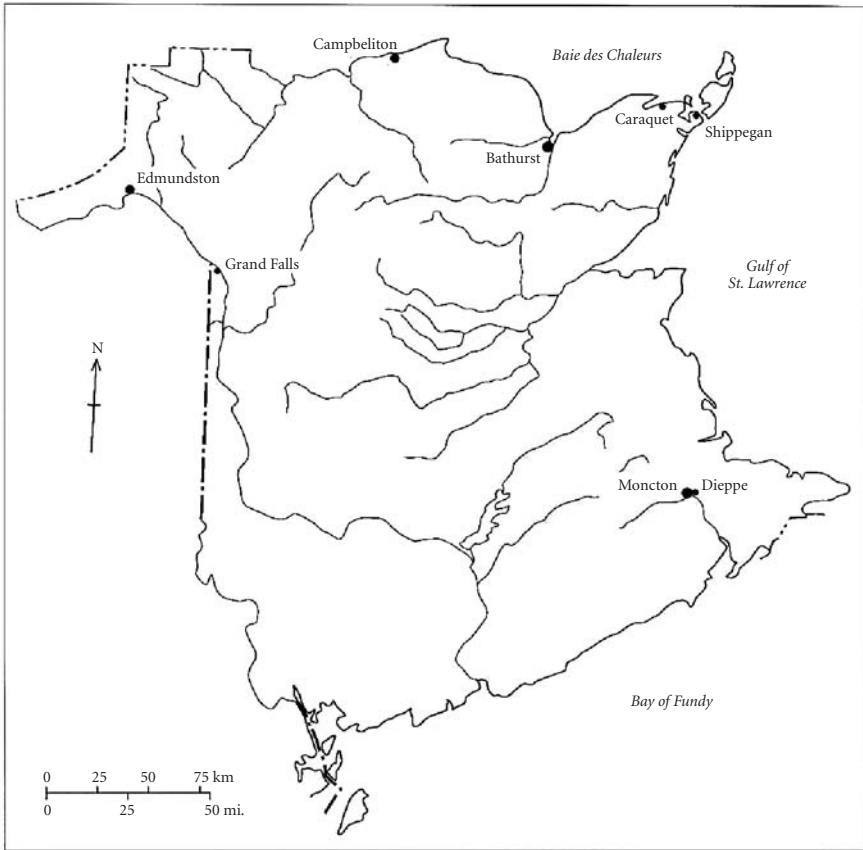
Map 5. New Brunswick Acadian Communities

six years' schooling used the Acadian plural at least 80% of the time.[12] As a first step in studying the variable, I compared overall agreement patterns in one of the most conservative of Acadian varieties, that of L'Anse-à-Canards, New-foundland, matching Flikeid and Péronnet's sample size and thus selecting 8 sociolinguistic interviews, in this case for four older and for younger speakers, and four female and four male speakers, for analysis. The results of this comparison are given in Table 2.[13]

The L'Anse-à-Canards speakers stand out as extremely conservative, far surpassing (at 99%) *-ont* the findings for the most conservative of the Nova Scotia communities.[14] Indeed, the three "non-Acadian" tokens found in the L'Anse-à-Canards corpus were two instances of *vont* (versus Acadian *allont*) and one of *font* (versus Acadian *faisont*).[15] Thus, when plurality was marked

Table 2. Proportion of usage of the traditional third person plural suffix *-ont* in seven Acadian communities*

Region	# of speakers	Average proportion of Acadian usage
Pubnico, Nova Scotia	8	.73
Chéticamp, N.S.	8	.84
Baie Ste-Marie, N.S.	6	.72
Île-Madame, N.S.	9	.78
Pomquet, N.S.	7	.87
Southeast New Brunswick	7	.70
L'Anse-à-Canards, NL	8	.99

* Nova Scotia and New Brunswick data from Flikeid and Péronnet (1989:228). Combined N for Nova Scotia and New Brunswick speakers = 2403; total N for Newfoundland speakers = 501.

on the verb, *-ont* was always used; what varied in these three cases was whether the Acadian stem was used or not.

6.   The case of subject relative clauses

Flikeid and Péronnet did find one clear case of statistically-significant inter-community variation in their study. Île-Madame and, to a lesser extent, Pomquet, both Nova Scotia communities, diverge from other local communities with respect to number marking in relative clauses, as shown in Table 3. What they found in this context was a far greater preponderance of default singular usage (i.e. plural agreement was not made) in one of the varieties, that of Île-Madame, and a greater preponderance of the Acadian third person plural usage, i.e., the *-ont* suffix, in the other varieties, with the exception of Pomquet, which falls in an intermediary position.[16] While Flikeid and Péronnet do not specify whether or not object relatives are involved, all of their examples consist of subject relatives. Examples (8) and (9) are taken from Flikeid and Péronnet (1989). Example (8) shows third person plural *veniont* and (9), the default singular, i.e. *vient*:

> (8) …des Anglaises   qui *veniont*   pis il   changiont
>   some English-women who were-coming and they were-changing
>   de classe   (205:104, Baie Ste-Marie)
>   of class
>   "…English women who came and they changed class"

Table 3. Comparison of the traditional third person plural suffix in relative clauses versus all other clause types: Nova Scotia and New Brunswick*

Region	Average proportion Acadian usage, relative clauses	Average proportion Acadian usage, other contexts
Pubnico, N.S.	.70	.73
Chéticamp, N.S.	.76	.84
Baie Ste-Marie, N.S.	.70	.72
Ile Madame, N.S.	.01	.78
Pomquet, N.S.	.61	.87
Southeast, N.B.	.64	.70

* All data from Flikeid and Péronnet (1989: 230).

Table 4. Comparison of the traditional third person plural suffix in subject relatives versus all other clause types: Newfoundland

Clause type	Plural marking	Singular marking	Ambiguous data
Subject relatives	8	30	27
All other clause types	423	1	12
Total Ns	431	31	39

(9) …avec des   grandes familles qui *vient*   alentour des
     with   some big   families who is-coming around   some
     maisons, là   (1:630, Île-Madame)
     houses   there
     "…with big families who come around houses"

Since the relative clause data were not part of their principal analysis, Flikeid and Péronnet do not attempt an explanation of these specific findings for Île-Madame and Pomquet.[17]

   Table 4 gives a parallel breakdown of data extracted from the Newfoundland subcorpus, some 501 tokens.[18]

   Examples of clear singular marking are given in (10)–(11). In (10) the Standard French plural would be *viennent* while the vernacular Acadian variant would be *venont*. In (11), the standard would be *ont* and the Acadian *avont* or *ont*. *Vient* in (10) and *a* in (11) are unambiguously singular verb forms.

(10) Il y a   ti   d'autres   histoires qui vous   *vient*
      there is INTERROGATIVE some other stories   that to-you comes
      dans l'idée?   (AC-2)
      in the idea
      "Are there other stories that come to mind?"

(11) Il y a   en masse de choses qu'*a*   arrivé?[19]   (AC-8)
      there are a lot   of things   that has   happened
      "Are there a lot of things that have happened?"

While the results in Table 4 show a striking difference between subject relatives
and other clause types with respect to number marking on the verb, they also
contain a fair number of ambiguous tokens. Tokens were classed as ambigu-
ous if they were homophonous with the regular standard or colloquial French
plural usage. Examples are given in (12)–(17)

(12) Il y a   un car devant nous-autres, un gros Chev, tu sais,   puis
      there is a car   in from of us   a big Chev   you know and
      deux vieilles femmmes qui [*drayve*]   (AC-8)
      two   old   women   who was?/were? driving
      "There's a car in front of us, a big Chev, you know, and two old women
      who were driving."

(13) Toujours, le Hood   avait été   coulé,   hein, comme ça, tous ceux
      anyway   the Hood had been flooded eh   like that   all those
      qu'étiont, qu'[*ete*]   à bord   ont été   noyé (self-correction)
      that were that was?/were? on board have been drowned.
      "Anyway, the Hood [name of boat] was flooded, eh, like that, all that were,
      that were on board were drowned."   (AC-3)

(14) Il y a   des choses   qui fait peur,   puis dáutres qu'est,
      there are some things which makes fear and some others
      qui [*dɔn*]   la joie   (AC-7)
      which is which gives?/give? the joy
      "There are things that make you afraid and there are others that are, that
      give you joy."

(15) Il y a   des   drôles d'affaires qui [*pas*]   dans la vie
      there are some strange things   which happens?/happen? in the life
      "There are strange things that happen in life."   (AC-2)

(16) C'était toujours dans la nuit qu'ils [*vwaye*]   des esprits
      it was always   in the night that they was?/were?-seeing some ghosts
      "It was always at night that they saw ghosts."   (AC-2)

(17) *Je* mangeais jusqu'à temps que les deux yeux [*vəne*]
      I was-eating until   time   that the two eyes was?/were?-coming
      tout blanc   (AC-7)
      all white
      "I ate until my two eyes became all white."

However, I argue that these potentially ambiguous cases are best analyzed as default singulars. The results presented in Table 2 show that, in all other contexts, Acadian -*ont* occurs in the vast majority of cases; indeed, there are no unambiguously -*ent* forms (such as *peuvent* "are able to" *or savent* "know," which involve plural stems). In theory, [drayve] in (12) could be *drivait* or *drivaient*. However, *drivaient* would seem unlikely, since, if the verb were in the plural, the expected ending would be the vernacular imperfect -*iont* rather than the standard -*aient* (recall from Section 5 that the plural form was always -*ont*). An examination of certain of the "ambiguous" sentences also points toward this analysis. Example (13) shows self-correction on the part of the informant, who has no unambiguously Standard French usage, so it seems most likely that the form [ete] is singular *était*, rather than plural *étaient*. (14) shows another "ambiguous" usage, the verb *donne* (homophonous with Standard French *donnent*), but it is contained within a string of unequivocal default singulars. Thus, both the overall quantitative results and the analysis of particular examples support reclassification of the ambiguous cases as default singulars.[20]

The data were coded for grammatical distinctions which might plausibly affect agreement. The hypothesis that clause type would turn out to be important was based on my familiarity with the L'Anse-à-Canards variety and with the Flikeid and Péronnet study. I initially made finer distinctions but collapsed them due to lack of variation among "other clause types." Constructing two "verb type" independent variables was a way of testing the hypothesis that, as Mougeon and Beniak (1991:92) note, morphological oppositions are best preserved in more frequent lexical items. A look at the frequency of individual verbs found that *avoir* "to have" (especially when used as an auxiliary) and modal verbs occurred often in the data. The data were then coded according to auxiliary versus main verb and according to simple versus modal verb. The data were also coded for the usual social variables age and sex of the speaker.

Some constructions occurred only rarely.[21] For example, there were only two instances of compound subjects (both with default singular verb forms) and only three *it*-clefts (all three containing default singular verb forms). Compound subjects were removed from the quantitative analysis while *it*-clefts were included under subject relatives.[22] Table 5 gives a breakdown of the results of multivariate analysis obtained through the use of the Goldvarb-2 program for Macintosh computers (Rand & Sankoff 1990).

These statistical results confirm that it is the relative construction which is at issue. Plural marking occurs in only 12% of tokens in this environment but in 97% of tokens in all other environments.[23] The examples below for object relatives show that it is subject relatives, not relative clauses in general, that are

Table 5. Results of multivariate analysis of factors potentially affecting third person plural agreement for L'Anse-à-Canards, Newfoundland

Factor groups	# pl. agr. made	Total	% pl. agr.	Factor weights
Clause type:				
Sub rel.	8	65	12%	.009
Other	423	436	97%	.670
Verb type:				
Auxiliary	85	103	83%	not sig.
Main	346	398	87%	
Verb type:				
Simple	417	483	86%	not sig.
Modal	14	18	78%	
Age:				
Younger	180	200	90%	not sig.
Older	251	301	83%	
Sex:				
Female	250	284	88%	not sig.
Male	181	217	83%	
TOTAL	431	501	86%	input = .941

relevant. As (18) and (19) show, in object relatives the verb agrees in number with a plural antecedent, as indicated by the presence of the *-ont* suffix.

(18)　Il y a　une gigue asteure qu'ils　*jouont*　　　　　(AC-4)
　　　there is a　jig　now　that they　play
　　　"There's a jig that they play."

(19)　…des maladies　que les bêtes　　　*avont* ou de quoi
　　　some　sicknesses that the farm-animals have　or　something
　　　comme ça　　　　　　　　　　　　　　　　　　　(AC-6)
　　　like　　that
　　　"…sicknesses that (farm) animals have or something like that."

How are we to explain the pattern shown in Table 5? Agreement in Standard French relative clauses works as it does in Standard English in that agreement ultimately holds between the verb and the head of the relative:

(20)　La fille　qui　*va*　à　l'école…
　　　the girl　who　goes　to the school
　　　"The girl who goes to school…"

(21)　Les filles　qui　*vont* à　l'école…
　　　the girls　who　go　to the school
　　　"The girls who go to school…"

In Standard English and French, the head of the relative and the verb contained within the relative must share the same number marking. This agreement in features is mediated by a coindexing relation between *who* or *qui* and the head of the relative. However, the Newfoundland equivalent, shown in (22), obviously cannot work in the same way.

(22)   Les filles qui *va*   à l'école…
       the girls who goes to the school
       "The girls who go to school…"

However, while number may not be transmitted to the verb in such cases, as (23) shows, gender **is** transmitted, since *les affaires* and the adjective *curieuse* are both feminine (the masculine form of the adjective would be *curieux*):

(23)   Les affaires qu'*a* été   curieuse…   (AC-4)
       the things   that have been strange
       "Things that were strange…"

Before we turn to a formal account of these facts, and to the puzzling question of how it is that gender can be realized but not number, we first turn to a special case of subject relatives, involving the expression *il y en a,* which will provide crucial evidence for the analysis.

7.   More subject relatives: The case of *il y en a…*

As was shown in Table 5, there were some instances of third plural agreement marking in subject relatives (8/73), in contradiction to the generalization. While one must expect some amount of data fluctuation, it is noteworthy that all of these instances of third person plural agreement have what appears at first glance to be just the clitic *en* as the head, i.e., they involve *il y en a* "there are some" or *il y en avait* "there were some". There is evidence that *il y en a* is a frozen expression in this variety and in a number of other varieties of colloquial French. Nadasdi (2000), for instance, makes such a case for Ontario French, pointing out that it is only in this expression that the clitic pronouns *y* and *en* actually co-occur in this variety. The same distribution is found in the L'Anse-à-Canards corpus. There is a clear contrast between examples such as (24) with bare *il y en a* and those such as (25) with *il y en a dix-huit* and (26) with *il y a deux hommes.*[24] The contrast centres on whether or not there is an identifiable head of the relative. Specifically, the relatives in (25) and (26) have identifiable heads while the one in (24) is essentially headless.

(24)   Des fois   il y en a   qui s'*assisont* sur les jambes…   (AC-4)
       some times they are some who REFL sit   on   the legs
       "Sometimes there are some who sit on their legs…"

(25)   Il y en a   dix-huit qu'*a* été tué   (AC-2)
       there are eighteen who has been killed
       "There were eighteen of them that were killed."

(26)   Il y a   deux hommes qu'*a*   venu à la porte   (AC-3)
       there are two   men   who has come to the door
       "There were two men who came to the door."

8. A formal account of third person plural marking

The agreement facts for the L'Anse-à-Canards variety of Acadian French to be
accounted for are as follows:

a.   In all clause types but subject relatives, overt plural agreement is virtually
     categorical (e.g. *les filles partont* "the girls are leaving").
b.   Subject relatives have default singular marking (e.g. *je connais les filles qui
     partit*) but overt gender marking is found (e.g. *les filles qu' était heureuse*
     "the girls that were happy").[25]
c.   The *il y en a* subtype of subject relatives in which *il y en a* has an overt head
     (e.g. *il y en a dix-huit…*) triggers default singular marking but bare *il y en
     a* has overt plural marking on the verb.

We begin with the status of relative *qui* (versus interrogative *qui*). In French
generally, interrogative *qui* is never used with inanimates but no such distinc-
tion obtains for relative *qui* (see for example (10) and (24)).[26] There is a long
tradition, dating from Kayne (1974), in French syntax to consider relative *qui*
as a variant of the *que* complementizer, more recently as the realization of *que* +
agr (Rizzi 1990), thus entailing a dependency between T and C. As for the syn-
tactic structure of the relative clause itself, the relative clause (a CP) is adjoined
to the NP as in (27):

(27)   DP[les NP[NP[filles] CP[qui partit]]]]

There has been considerable debate dating from the 1980s onwards as to
whether the head of relative clauses is base-generated outside CP and linked to
a null operator in CP by a predication relation or whether it raises to SpecCP.
Since neither approach appears to have implications for the present analysis, I

adopt the former, as exemplified in (28) where, for clarity of exposition, I have inserted the overt morphology of the complementizer and the verb:[27]

(28) $_{DP}$[les $_{NP}$[$_{NP}$[filles]] $_{CP}$[OP$_i$ qui$_j$ $_{TP}$[t$_i$ t$_j$ partit$_k$ $_{VP}$ [t$_i$ t$_k$]]]]

The patterns found in a-c above show variable number marking on the verb depending on the construction type. Instead of adhering to early Minimalism (Chomsky 1993), whereby verbs come fully inflected in the lexicon, I assume, following Halle and Marantz (1993) and the theory of Distributed Morphology, that a linguistic item is a bundle of morphological features. In this view, inflectional elements are inserted on the verb in the phonological component, following the syntax. Thus, in the analysis which follows, it is the abstract features for person, gender and number (the so-called Phi-features) which are checked in the syntax, not the overt morphology.

Under Minimalism, the operation 'Agree' establishes a checking relation between a syntactic object and some other syntactic object. The Phi-features person, gender and number, the relevant features for the current analysis, are interpretable (by the semantics) on the DP (e.g. *les filles* designates a set of individuals) but not on the verb (e.g. there is no plurality of events in *les filles partont*). Uninterpretable features must be checked through the operation Agree: for instance, the head Tense must be paired with the plural feature on the subject DP, thus associating plurality with Tense; Tense itself must be paired with the verb, thus associating plurality with the verb. (Tense and the subject DP also agree for (nominative) case.)
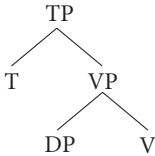
Subsequently, the DP subject moves to the specifier position of TP and the verb raises to T. Once the phase is complete (Chomsky 2001), the derivation undergoes Spell-Out and is submitted to the phonological and semantic components. The checked features are accessible to the phonological component and it is here that overt agreement marking (in this case the third person plural -*ont* suffix) is inserted on the verb. The L'Anse-à-Canards data also show overt gender agreement with predicate adjectives. Gender morphology is likewise inserted in the phonological component (e.g. *les filles sont heureuse* "the girls are happy"). This, then, accounts for the pattern noted in a straightforward fashion.

As for the patterns found in b and c, the analysis is necessarily more complex. The discussion of subject relatives and the status of the *qui* (<*que*) complementizer leads us to assume multiple Agree relations in the derivation of these structures. Recall that Rizzi (1990) has argued that relative *qui* is the overt realization of *que* + agr: it would follow logically that agr in C is co-indexed with agr in T. Further, Pesetsky and Torrego (2000: fn.73) suggest that, like English *that*, French *qui* is an instance of T moved to C.[28] I assume, then, that subject relatives have the following structure:

```
                    DP
                  /    \
                DP       CP
                |       /   \
               agr   OP_i     C'
                     |       /   \
                    agr     C      TP
                           / \    /  \
                   que_j + agr  T      ...
                               / \
                             t_j  agr
```
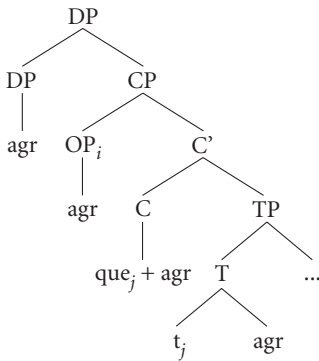
Feature checking involves (a) for Tense, checking both *u*Case and *u*Phi-features (where *u* = uninterpretable) against OP, the null operator; (b) for the null operator, checking *u*Case against Tense and checking *u*Rel against C (C has an interpretable Rel feature); and (c) for the complementizer, checking *u*Tense via T-to-C movement (Tense bears Phi-features from *que* + agr = *qui*), and checking *u*OP against the null operator.

While we have accounted for (i.e. checked) all of the relevant uninterpretable features in the structure, we have not addressed the issue of how it is that plural morphology is sometimes overt, sometimes not. The basic pattern is when there is an overt "antecedent" and when there is also a predication relationship such as obtains between an NP and a relative clause, plural is not overtly marked but rather there is default singular marking. Since number marking is variable in this way, some of the phonological features of a lexical item must be available from the beginning of the derivation: i.e., the head of the relative must carry some phonological features or it would not otherwise be possible for the phonological component to "decide" that agreement morphology should not be spelled out on the verb. Within the theory of Distributed Morphology, it is not possible to change the feature specification of a lexical item, but it is possible to insert underspecified (default) morphemes.

In the presence of a phonologically visible head (e.g. *les filles* "the girls"), then, default singular marking is inserted in the phonological component.

This is the case with subject relatives of the non-*il y en a* type, where we do not get overt plural agreement with an antecedent but rather default number marking. While the plural feature is not spelled out on the verb, the adjective does overtly agree in gender with the head of the subject relative clause, as in (29):[29]

(29) $_{DP}$[les $_{NP}$[$_{NP}$[filles]] $_{CP}$[OP$_i$ qui$_j$ $_{TP}$[t$_i$ t$_j$ a été$_k$ $_{VP}$ [t$_i$ t$_k$ $_{SC}$[t$_i$ curieuse]]]]]

Likewise, in *il y en a* relatives with an identifiable head (e.g. *il y en a dix-huit* "there are eighteen of them"), there is no insertion of the *-ont* suffix but rather the verb always occurs in the default singular form. In contrast, with bare *il y en a*, there is variation in individual speakers' usage. For some speakers, constructions with *il y en a* pattern like other subject relatives; for others there is overt plural marking.[30] What would account for plural marking with bare *il y en a*? The most likely explanation is that overt plural marking on the verb in what is a headless construction for these speakers is a strategy to spell-out the number of the subject in the absence of an overt head bearing this feature. We end up, then, with a unified account of the data which allows that agreement happens in all cases (both number agreement and, where applicable, gender agreement, i.e. the Phi-features all behave in the same way). What varies is whether or not the plural agreement maker is spelled out on the verb or not. In subject relatives, if plural agreement is recoverable, the phonological component will not insert an agreement morpheme; if it is not recoverable, an agreement morpheme will be inserted.

9.   Comparison with other varieties of French

While most studies of subject-verb agreement in French do not go into the degree of detail presented above, there is clear evidence of some degree of variation along similar lines. In a study of third person plural marking in Prince Edward Island French, King and Nadasdi (1996) found a contrast between speakers who were resident of Abram-Village and those residing in Saint-Louis. In the former, there is a fairly high degree of contact with external varieties of French and a French medium educational system in place since the early 1960s, whereas the latter community has little contact with other varieties (Map 6).

The Prince Edward Island case presents a more complex picture than did the Newfoundland one, due first of all to a wider array of variants, including,
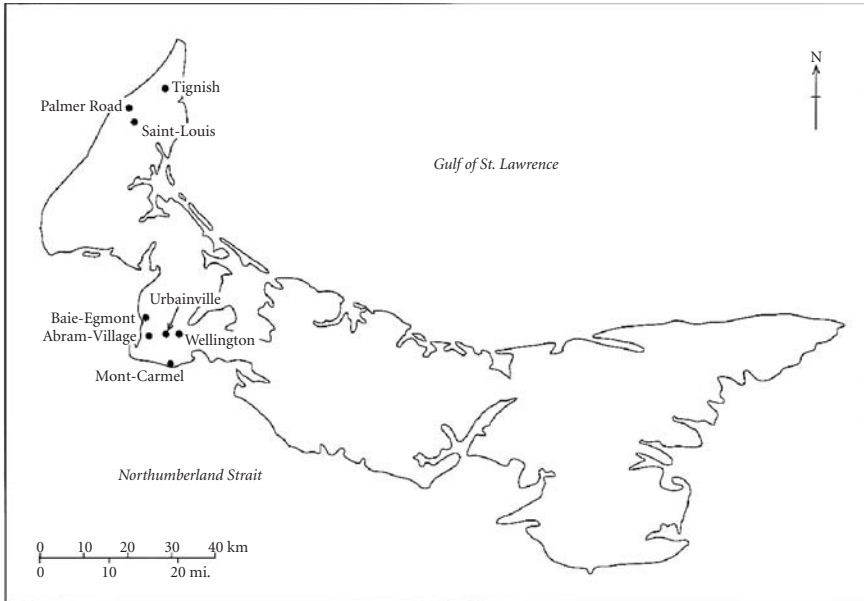
Map 6. Prince Edward Island Acadian Communities

along with clearly normative French usage as well as vernacular usage, a much higher proportion of ambiguous cases. Therefore a larger sample size was used for this particular study, involving 5000 tokens for 26 speakers, representative of the two communities, both sexes and a wide age range. The overall results according to community are presented in Table 6.

The fact that 'community' emerged as significant in multivariate analysis of the potential effects of a number of linguistic and social variables, is explicable in terms of the differing social circumstances of French in the two locales.[31] With regard to linguistic constraints on variation, however, the clause type constraint, which approaches categoriality in Newfoundland, did exert some effect on variation: however, while subject relative clauses disfavour overt plural marking on the verb, the effect was fairly weak.[32]

Table 6. Comparison of the traditional third person plural suffix by community: Prince Edward Island

Community	# of 3 ppl tokens	# of *-ont* tokens	Proportion of *-ont* usage
Saint-Louis	1733	1432	.83
Abram-Village	3109	2105	.68

In their discussions of working-class European French, both Frei (1929) and Bauche (1946) give examples of default singulars in subject relative clauses:

(30) J' aime pas les femmes qui *boit*   (from Frei 1929)
       I like not the women who drinks
       "I don't like women who drink."

(31) Les hommes qu' *a* vendu la France…   (from Bauche 1946)
       the men who has sold the France
       "The men who sold France…"

The actual status of default singulars is not made clear in these works, though we may assume their use to be sporadic. Such usage is also the case for present-day Ontario French, where Mougeon and Beniak (1995) have linked certain types of non-agreement (as in (32), with a main clause lexical subject) to English dominance but have found that French-dominant speakers only have default singulars in subject relatives, as in (33), albeit sporadically.

(32) *Les* singes *peut* faire qu' est-ce qu' on peut faire
       the monkeys can to-do what it is that we can to-do
       "Monkeys can do what we can do."

(33) Y a beaucoup de choses qui s'*produit*
       there are a lot of things that REFL go-on
       "There are a lot of things going on."

If there is only a slight tendency towards default singulars in subject relatives in these other varieties of French, one might ask how two varieties of Acadian French became so advanced, i.e. those of Île-Madame, Nova Scotia and L'Anse-à-Canards, Newfoundland. The fact that several Acadian varieties behave no differently from other varieties of French points to New World and most likely post-Deportation innovation. Following their return from exile dating from the mid to late 1700s, the Acadian people formed small, isolated communities, some of which underwent fairly rapid assimilation to English while others remained Francophone enclaves. In Newfoundland, settlement was fairly late with the first Acadian settlers of western Newfoundland arriving from Chéticamp, Nova Scotia near the end of the 18th century. There has been no contact of which I am aware between the Newfoundland settlers and the residents of Île-Madame. As shown in Table 3, the subject relative vs. all other clause types split does not occur in the French of Chéticamp while the distinction is clearcut in the Newfoundland variety and in the French of Île-Madame. While Chéticamp speakers exhibit weak tendencies towards the agreement pattern discussed here (76% -*ont* in relatives, 84% -*ont* in other clause types), the

oldest French Newfoundlanders we have recorded (people born at the turn of the century) have near-categorical use of default singulars in subject relatives. This would then point to an independent development whereby a tendency towards a default singular in this construction has become a fully-fledged feature of the grammar. The two Acadian varieties with default singular usage as the norm have a common sociolinguistic feature: there is very low, indeed almost nonexistent, normative pressure in the direction of Standard French. This would, likewise, appear to be the case for Cajun French, where recent research (Dubois, King, & Nadasdi 2004) has shown that while the traditional variant is much less frequent than in the Acadian varieties discussed above, in subject relative relative clauses default singulars are categorical.[33] This situation provides a setting in which rapid change in the direction described here could occur.

However, it is not the case that such change necessarily occurs in such a setting, as evidenced by the Saint-Louis, Prince Edward Island facts, where clause type exhibits a weak effect on variation. This result is somewhat surprising, given that Saint-Louis residents have stood out as heavy vernacular users in a number of studies of other variables, such as *je... ons* retention (King, Nadasdi, & Butler 2004) and as highly innovative code-switchers (King & Nadasdi 1999). Thus, a situation of such low normative pressure may provide a setting in which the default singular pattern may take hold, but it is not necessarily the case that this will actually happen. In other words, it may provide a necessary but not a sufficient condition for change.

10. Conclusion

This paper has attempted to account for variation in one aspect of the morphosyntax of Acadian French, third person plural marking, along both social and linguistic dimensions. On the one hand, it has involved quantitative methodology and comparative sociolinguistic (see Tagliamonte 2002) analysis; on the other, it has appealed to formal theory for an understanding of finely-grained linguistic variation. The comparative sociolinguistic perspective has elucidated the social factors which allow the loss of plural marking in subject relative clauses and its retention in other clause types, specifically, lack of exposure to agreement patterns from other varieties of French and concomitant low normative pressure. Within the subject relative construction, variationist methodology has picked out a rather subtle point of variation, the *il y en a* subtype, variation which might well go unobserved outside the analysis of a fairly large corpus of data. While the sociolinguistic analysis has allowed

the identification of variation in agreement marking, the structural analysis has provided the architecture for allowing default singulars, thus providing a principled account of variation in individual speakers' grammars.

Notes

1. Presumably speakers of Standard English who are bidialectal can get both readings.

2. Labov's own figures for the journal are for the period 1989–1993, to which I have added data for 1994–2003. I calculated the proportion of articles devoted to grammatical variation (morphology and syntax) versus those devoted to phonological variation and found that the former slightly outnumbered the latter, with 52% versus 48% respectively (articles on discourse, semantics, lexicon, etc. were not considered).

3. King (2000) provides a sociohistorical overview and a grammatical sketch of Acadian French.

4. All but sporadic contact with English dates from the 20th century, 1935 specifically, and the government-sponsored relocation of English Newfoundlanders of (southeast) Anglo-Irish descent from the province's south coast to a newly-created Anglophone community, Lourdes, on the Port-au-Port peninsula (Butler 1995:46). For a discussion of the English origins of settlers to Atlantic Canada more generally, see Edwards (1998).

5. Only data from fully fluent French speakers are used in the analyses presented here.

6. I follow convention here and write first and third person plural endings as *-ons* and *-ont* respectively. It is not certain whether *-ont* arose through analogy with the first person plural, or with common third person plural forms *sont* (third person plural of *être* "to be"), *vont* (*aller* "to go") and *font* (*faire* "to do"). Table 1 also shows, in the Perfect column, variation in the form of the auxiliary *avoir* "to have," with the first person singular *ai/as*, the first person plural *avons/ons*, and the third person plural *avont/ont*. In this variety, all are vernacular variants, i.e. they do not involve style shifting to more standard usage.

7. This turns out to be an important point, since, as noted above, the Newfoundland communities have mixed settlement: they owe a minority of their Francophone settlers to 19th century immigrants from France and the majority to Acadians from the Chéticamp area of Nova Scotia. Since the metropolitan French colonists came from Brittany, both groups would have had the traditional third person plural variant.

8. The third person singular and plural pronominal forms *il* and *ils* tend to be homophonous in colloquial French. This is the case for Newfoundland Acadian, where *ils* is (to my knowledge) never realized with [z] in liaison contexts (e.g. *ils arrivont* "they arrive" [ilar-ivō]). Similarly, *ils* with a [z] of liaison is at least unusual in Prince Edward Island French and in New Brunswick and Nova Scotia varieties as well, as evidenced by Flikeid and Péronnet's (1989) orthographic choice *il* for the third person plural pronoun, as in (8) below. Note as well that, as is often the case in colloquial French generally, *ils* is unmarked for gender; the Standard French third person feminine plural *elles* is unattested in our corpora.

9. Acadian French has so-called Doubly-filled COMP (King 1991).

10. With the exception of two verbs, *mourir* "to die" and *naître* "to be born", Acadian French almost invariably has only *avoir* "to have" as an auxiliary (King & Nadasdi 2001). In the Standard French equivalent of (7), *venir* "to come" would be conjugated with *être* "to be".

11. Unfortunately, while Flikeid and Péronnet say there are a total of 2,403 third person plural tokens (p. 222) in their corpus, they report only proportions in their quantitative results, which represents an incomplete picture of variation. Since the methodology of Flikeid's Nova Scotia study (Flikeid 1994) was very similar to my own, I hypothesize that their findings here are based on about the same amount of data per individual as was my study, where no speaker contributed fewer than 45 third person plural tokens. That is, I would not expect the number of third person plural tokens to vary so dramatically from sociolinguistic interview to sociolinguistic interview that a particular individual contributed little data (and that individual to be retained in the sample). However, some degree of caution is needed in interpreting Flikeid and Péronnet's results.

12. A number of recent sociolinguistic studies suggest that education is becoming an increasingly important factor, more important than social class. See for example Tagliamonte (1998) for a study of grammatical variation in a British variety, that of York.

13. The general results presented here are taken from King (1994). I thank Cambridge University Press for permission to reproduce portions of that early paper.

14. Though the quantitative results reported on here deal with only 8 speakers, subsequent work on third person plural agreement (e.g. King & Nadasdi 1996) reveals that these speakers' usage is representative of usage in Butler's approximately 1,000,000 word L'Anse-à-Canards corpus more generally.

15. Such usage (i.e. third person plural verb forms *vont* and *font*) overlaps with Standard French. However, given that such forms occur in a number of varieties where exposure to the standard through schooling ranges from heavy to nonexistent, I refrain from labelling them as standard since they also occur in colloquial French.

16. While I use the term default singular, subject/verb non-concord is more common terminology in the sociolinguistics literature (Chambers 1995:243). Variable agreement phenomena have been the subject of considerable variationist research, such as the much-studied case of the so-called "Northern Subject Rule" for English, whereby a number of varieties of English tend to have agreement with pronominal subjects but not necessarily with a full noun phrase (see for example Adger and Smith (this volume), Montgomery (1994), and Tagliamonte (2002). Henry (2002:278) uses the term "singular concord" for cases like "The books goes on the shelf."

17. However, in their general discussion of intercommunity differences they note a less firmly entrenched tradition of French education and less contact with other Francophones in these two communities than in the other ones.

18. As noted earlier with regard to third person plural more generally, these results are in line with what was subsequently found in the larger L'Anse-à-Canards corpus. Indeed, though not subject to previous quantitative analysis, lack of plural agreement in subject relatives has been the subject of commentary since the first linguistic work on Newfoundland Acadian French (King 1978).

19. Relative *qui* reduces to *qu'* [k] in the environment before a vowel, thus patterning with the *que* complementizer (e.g. *Je sais qu'il a parti* "I know that he left"). We shall see below that relative *qui* is best regarded as a variant of the *que* complementizer.

20. It may seem controversial that the ambiguous forms were reclassified as singulars rather than being removed from the data as is more typical of variationist analysis. I take the (pragmatic) approach that this is a linguistically-informed decision which bolsters the numbers for quantitative analysis. A comparison of the "ambiguous" tokens with the overtly-marked singular ones reveals no skewing according to potential conditioning factors (e.g. there were no observable differences in distribution across clause type).

21. Cornips and Corrigan (2005) discuss in detail the problem of low frequency of occurrence in corpora for particular types of (morpho)syntactic variables, along with some ways around it.

22. While it is not clear if *it*-clefts are frozen expressions or are indeed generated by the grammar, their influence on these results is obviously minimal.

23. I return to that small number of cases where plural agreement is actually made in subject relatives below.

24. As there were only twelve instances of *il y en a* initially considered here, additional data for these speakers and for several other residents of L'Anse-à-Canards were added to the present analysis of *il y en a* structures. The new speakers, like the initial group, were selected so as to vary by age and sex. Like the other eight speakers, none had been educated in French, none had achieved more than secondary education and none worked outside the community.

25. In Newfoundland Acadian, the third person singular of the verb *partir* "to leave" in the present tense is the regularized form *partit*, not Standard French *part*.

26. Newfoundland Acadian provides additional evidence of a distinction between the two in that interrogative *qui* and relative *qui* differ in their phonetic realization. Interrogative *qui* undergoes the typically Acadian phonological process whereby velar stops palatalize to become affricates before non-low front vowels; relative *qui* does not undergo this process. To my knowledge this distinction obtains in all Acadian varieties which still have palatalization of velar stops.

27. Under this analysis, OP_i is a null operator, a silent WH-word, which is linked to the head of the relative, *les filles*, by an interpretive (predication) relation. The same sort of analysis would be given to the English equivalent *the girls that left*.

28. Under the Pesetsky and Torrego analysis, it is unclear why T moves to C but the verb (which in French raises to T) does not.

29. SC stands for small clause.

30. A look at individual speakers' usage shows that, while not in all cases categorical, individuals strongly favour one option or another.

31. King and Nadasdi (1996) presents complete results for Prince Edward Island. This study considers, along with community, a number of other social factors, including level of French language education and position in the linguistic marketplace (Sankoff & Laberge 1978).

32. Another conditioning effect emerged in this study, not uncovered in previous studies of the variable, i.e., the effect of subject type. Note that in our investigation of linguistic constraints on variation, a somewhat different partitioning of third person plural variation was used in this study, in that the data were analysed along a singular vs. plural dimension (i.e. all overt manifestations of plurality, not uniquely vernacular usage, are contrasted with cases of overt singular marking). With the pronominal subject *ils* plural marking was more likely to occur than with a lexical subject, 79% versus 71% respectively, which was found to be statistically significant in the multivariate analysis. Subsequent studies of another Acadian variety (Beaulieu & Cichocki 2003 for north-eastern New Brunswick Acadian) and a close relative, Cajun, (Dubois, King, & Nadasdi 2004) also found evidence of such an effect. Intuitively, these results seem related to the fact that *ils* is not overtly marked for number, being homophonous with third person masculine singular *il* "he". Thus when there is a subject pronoun, number cannot be recovered from the subject, similar to the bare *il y en a* case noted above for the Newfoundland variety. But more would appear to be at issue since, crosslinguistically, pronouns are more likely to trigger agreement than are other types of subjects (Henry 2002:277). An explicit account is beyond the scope of this paper, beyond the observation that variation linked to subject type would appear to fall out from universal tendencies. It should be noted, though, that contact with English cannot be argued to play a role here (see Note 15 and the Northern Subject Rule) since the subject type constraint holds for the northeastern New Brunswick case, where contact with English is minimal (much less than in L'Anse-à-Canards) as it does for the Cajun one, where contact is intense.

33. It may be argued that loss of the *-ont* suffix is part of the general breakdown of verb morphology in Cajun; if this is the case, it is still noteworthy that the change is most pronounced, indeed gone to completion, in subject relatives.

References

Arsenault, Georges (1987). *Les Acadiens de l'Île*. Moncton: Éditions de l'Acadie.

Auger, Julie (1994). Pronominal Clitics in Colloquial French: A Morphological Analysis. Ph.D. dissertation, University of Pennsylvania.

Auger, Julie (1998). "Le redoublement des sujets en français informel québécois: Une approche variationiste". *Canadian Journal of Linguistics, 43*(1), 37–63.

Bauche, Henri (1946). *Le langage populaire* 4. Paris: Payot.

Beaulieu, Louise & Patricia Balcom (1998). "Le statut des pronoms personnels sujets en français acadien du nord-est du nouveau-Brunswick". *Linguistica Atlantica, 20*, 1–27.

Beaulieu, Louise & Wladyslav Cichocki (2003). "Grammaticalisation et pertes des marques d'accord sujet-verb en français acadien du nord-est du nouveau-Brunswick". In Sandra Clarke (Ed.), *Papers from the 26th annual meeting of the Atlantic Provinces Linguistic Association* (pp. 121–143). St. John's, NF: Memorial University.

Brunot, Ferdinand (1967). *Histoire de la langue française.* 15 vol. Paris: Librairie Armand Colin.

Butler, Gary R. (1991). *Saying isn't Believing: Conversation, Narrative and the Discourse of Belief in a French Newfoundland Community*. St. John's, NL: Institute for Social and Economic Research and Philadelphia: Publications of the American Dialect Society.

Butler, Gary R. (1995). *Histoire et traditions orales des franco-acadiens de Terre-Neuve*. Quebec: Septentrion.

Chambers, J. K. (1995). *Sociolinguistic Theory*. Oxford: Blackwell.

Chomsky, Noam (1993). "A Minimalist Program for Linguistic Theory". In Kenneth Hale & Samuel Jay Keyser (Eds.), *The View from Building 20* (pp. 1–52). Cambridge, MA: MIT Press.

Chomsky, Noam (2001). "Derivation by Phase." In Michael Kenstowicz (Ed.), *Ken Hale: A Life in Language* (pp. 1–52). Cambridge, MA: MIT Press.

Cornips, Leonie & Karen Corrigan (2005). "Convergence and Divergence in Grammar". In Peter Auer, Frans Hinskens, & Paul Kerswill (Eds.), *Dialect Change: Convergence and Divergence in European Languages* (pp. 96–134). Cambridge: Cambridge University Press.

Dubois, Sylvie, Ruth King, & Terry Nadasdi (2004). "Past and Present Agreements: A Comparison of Third Person Plural Marking in Acadian and Cajun French". Paper presented at *Sociolinguistic Symposium 15*, Newcastle.

Edwards, John (Ed.). (1998). *Language in Canada*. Cambridge: Cambridge University Press.

Flikeid, Karin (1989). "Moitié anglais, moitié français?: Emprunts et alternance de langues dans les communautés acadiennes de la Nouvelle Écosse". *Revue québécoise et linguistique théorique et appliquée, 8*(2), 177–227.

Flikeid, Karin (1994). "Origines et évolution du français acadien à la lumière de l'étude de la diversité contemporaine". In Raymond Mougeon & Édouard Beniak (Eds.), *Les origines du français québécois* (pp. 275–236). Québec: Presses de l'Université Laval.

Flikeid, Karin & Louise Péronnet (1989). "N'est-ce pas vrai qu'il faut dire 'j'avons été'?: Divergences régionales en acadien". *Français moderne, 57*(3/4), 219–228.

Frei, Henri (1971). *La grammaire des fautes*. Geneva: Slatkine Reprints. (Original published 1929 by Paris-Genève.)

Halle, Morris & Alec Marantz (1993). "Distributed Morphology and the Pieces of Inflection". In Kenneth Hale & Samuel Jay Keyser (Eds.), *The View from Building 20* (pp. 111–176). Cambridge, MA: MIT Press.

Heap, David (2001). "Split Subject Pronoun Paradigms: Feature Geometry and Under-specification." In Teresa Satterfield, Christina Tortora, & Diana Crest (Eds.), *Current Issues in Linguistic Theory: Selected Papers from the XXIXth Linguistic Symposium on Romance Languages* (pp. 124–139). Amsterdam & Philadelphia: John Benjamins.

Henry, Alison (2002). "Variation and Syntactic Theory". In J. K. Chambers, Peter Trudgill, & Nathalie Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 267–282). Oxford: Blackwell.

Kayne, Richard (1974). "French Relative Que". In Marta Luján & Fritz Hensey (Eds.), *Current Studies in Romance Linguistics* (pp. 255–299). Washington: Georgetown University Press.

King, Ruth (1978). Étude phonologique et morphologique du parler français de l'Anse-à-Canards, Terre-Neuve. MA thesis, Memorial University of Newfoundland.

King, Ruth (1991). "WH-Words, WH-Questions, and Relative Clauses in Prince Edward Island French". *Canadian Journal of Linguistics, 36*(1), 65–85.

King, Ruth (1994). "Subject-Verb Agreement in Newfoundland French". *Language Variation and Change, 6*, 239–253.

King, Ruth (2000). *The Lexical Basis of Grammatical Borrowing: A Prince Edward Island French Case Study*. Amsterdam & Philadelphia: John Benjamins.

King, Ruth & Gary R. Butler (in press). "Les franco-terreneuviens et le franco-terre-neuvien". In Julie Auger & Albert Valdman (Eds.), *Le français en amérique du nord*. Québec: Presses de l'Université Laval.

King, Ruth, France Martineau, & Raymond Mougeon (2004). "L'Usage des pronoms personnels sujet inclusifs de la 4e personne en français populaire hexagonal, québécois et acadien: Recherche sociolinguistique diachronique". Paper presented at the *Conference grammaire comparée des variétés de français d'amérique*, Avignon.

King, Ruth & Terry Nadasdi (1996). "Sorting Out Morphosyntactic Variation in Acadian French: The Importance of the Linguistic Marketplace". In Jennifer Arnold et al. (Eds.), *Sociolinguistic Variation: Data, Theory and Method. Selected Papers from NWAVE-23* (pp. 113–128). Stanford: Centre for the Study of Language and Information (CSLI).

King, Ruth & Terry Nadasdi (1997). "Left Dislocation, Number Marking and Non-Standard French". *Probus, 9*, 267–284.

King, Ruth & Terry Nadasdi (1999). "The Expression of Evidentiality in French-English Bilingual Discourse". *Language in Society, 28*(3), 355–366.

King, Ruth & Terry Nadasdi (2001). "How Auxiliaries Be/Have in Acadian French". In Gisèle Chevalier (Ed.), *Papers from the Annual Meeting of the Atlantic Provinces Linguistic Association* (pp. 61–72). Moncton: Université de Moncton.

King, Ruth, Terry Nadasdi, & Gary R. Butler (2004). "First Person Plural in Prince Edward Island Acadian French: The Fate of the Vernacular Variant *je… ons*". *Language Variation and Change, 16*(3), 237–255.

Kroch, Anthony (2000). "Syntactic Change". In Mark Baltin & Chris Collins (Eds.), *The Handbook of Contemporary Syntactic Theory* (pp. 629–739).

Labov, William (1972). *Language in the Inner City*. Philadelphia: University of Pennsylvania Press.

Labov, William (1994). "Analyzing Variation Above the Level of Phonology". Panel presentation at *NWAV-23*, Stanford University.

Meechan, Marjory & Michele Foley (1994). "On Resolving Disagreement: Linguistic Theory and Variation – *There's Bridges*". *Language Variation and Change, 5*, 113–139.

Meyerhoff, Miriam (2000). "The Emergence of Creole Subject-Verb Agreement and the Licensing of Null Subjects". *Language Variation and Change, 12*, 203–230.

Montgomery, Michael (1994). "The Evolution of Verb Concord in Scots". In A. Fenton & D. Macdonald (Eds.), Studies in Scots and Gaelic: *Proceedings of the Third International Conference on the Languages of Scotland* (pp. 81–95). Edinburgh: Cannongate Academic.

Mougeon, Raymond & Édouard Beniak (1991). *Linguistic Consequences of Language Contact and Restriction: The Case of French in Ontario, Canada*. Oxford: Oxford University Press.

Mougeon, Raymond & Édouard Beniak (1994). "Le non-accord en nombre entre sujet et verbe en français ontarien: Un cas de simplification?" *Présence francophone, 46*, 53–63.

Nadasdi, Terry (1995). "Subject NP Doubling, Matching and Minority French". *Language Variation and Change, 7*(1), 1–14.

Nadasdi, Terry (2000). *Variation morphosyntaxique et langue minoritaire: Le cas du français ontarien*. München: LINCOM Europa.

Nyrop, Kristoffer (1979). *Grammaire historique de la langue française*. 6 vol. Geneva: Slatkine Reprints. (Original published 1889–1936).

Pesetsky, David & Esther Torrego (2000). "T-to-C Movement: Causes and Consequences". In Michael Kenstowicz (Ed.), *Ken Hale: A Life in Language*. Cambridge, MA: MIT Press.

Poplack, Shana & Sali Tagliamonte (2001). *African American English in the Diaspora*. Malden, MA: Blackwell Publishers.

Rand, David & David Sankoff (1990). *Goldvarb: A Variable Rule Application for the Macintosh*. Montreal: Centre de Recherches Mathématiques.

Rizzi, Luigi (1990). *Relativized Minimality*. Cambridge, MA: MIT Press.

Ross, Sally & Alphonse Deveau (1992). *The Acadians of Nova Scotia: Past and Present*. Halifax: Nimbus.

Sankoff, David (1988). "Sociolinguistics and Syntactic Variation". In Frederick J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey. Vol. IV. Language: The Socio-Cultural Context* (pp. 140–161). Cambridge: Cambridge University Press.

Sankoff, David & Suzanne Laberge (1978). "The Linguistic Market and the Statistical Explanation of Variability". In David Sankoff (Ed.), *Linguistic Variation: Models and Methods* (pp. 239–250). New York: Academic Press.

Sankoff, Gillian (1990). "The Grammaticalization of Tense and Aspect in Tok Pisin and Sranan". *Language Variation and Change, 2*, 295–312.

Tagliamonte, Sali (1998). "*Was/Were* Variation Across the Generations: View From the City of York". *Language Variation and Change, 10*, 153–191.

Tagliamonte, Sali (2002). "Comparative Sociolinguistics". In J. K. Chambers, Peter Trudgill, & Nathalie Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 729–763). Oxford: Blackwell.

Wilson, John & Alison Henry (1998). "Parameter Setting Within a Socially Realistic Linguistics". *Language in Society, 5*, 141–187.

Syntactic variability across georgraphical space

Word order variation in three-verb clusters and the division of labour between generative linguistics and sociolinguistics⋆

Sjef Barbiers

Meertens Instituut, Royal Netherlands Academy of Sciences

1. Introduction

In Standard Dutch, the order of verbs in clause final three-verb clusters is 1–2–3 (1a). In Standard Frisian, the order is 3–2–1 (1b)[1]

(1) a. dat hij moet₁ kunnen₂ zwemmen₃   Standard Dutch
       that he must can.INF swim.INF
       "that he must be able to swim."
   b. dat er de doar grien fervje₃   kinne₂ woe₁ Standard Frisian
       that he the door green paint.INF can.INF wanted
       "that he wanted to be able to paint the door green."

When non-standard varieties in the Dutch language area, including Friesland and Belgium are taken into account we find a bewildering variation (cf. Stroop 1970; Evers 1975, among many others). Five of the six logically possible word orders in three-verb clusters are attested. The only order that appears to be systematically absent is 2–1–3. Part of the distribution of the different orders seems to be determined geographically. In addition, we find speaker-dependent variation. There are many speakers who accept two or more of the orders, up to four. A full account of this variation requires both a generative and a sociolinguistic approach. The main goal of this paper is to establish the division of labour between these two approaches.

More specifically, I propose a generative analysis that explains the attested variation space and then formulate a number of questions that generative theory cannot explain and that require sociolinguistic research to be answered

satisfactorily. I argue that variation and optionality are an inherent property of grammatical systems. Individual speakers and communities pick their choice from the options provided by their grammatical system, but they never pick beyond these options.

Multiple variation as found in verb clusters poses a problem for those versions of generative theory that assume there to be no variation or optionality in I-language (cf. Chomsky 1995). The literature on parametrisation provides at least two ways out. The first way is to say that when a speaker has two options, he has two grammars (cf. Kroch 1994). This would require four grammars for certain speakers of Dutch dialects to capture the variation in verb clusters, and an explosion of grammars if other cases of optionality such as PP Extraposition and scrambling are taken into account.[2] An alternative explanation is to assume that every speaker allows only one order, but that their judgements on the other orders are determined by what sounds familiar, i.e. word orders from surrounding dialects (e.g., Zuckerman 2001). Although this solution cannot be rejected out of hand either, there does not seem to be a simple relation between geographic distance and familiarity with other dialects. To test this solution, the individual familiarity with dialects other than one's own should be measured, a rather complicated task.

A problem for both types of solution is that they require accounts involving fine-grained parametrisation. One of the reasons why the Principles and Parameters framework (Chomsky 1981 and subsequent work) was replaced by the Minimalist Program (Chomsky 1995) was that the number of parameters had become so large that parameter theory had become virtually meaningless (cf. Baker 1996; Henry 2002). Another problem is the hypothesis that I-languages (internalised language specific grammars) do not show internal optionality, in other words, variation is unexpected if language structures are biological structures, since it is clear that the properties of biological structures of a particular type may vary within certain limits.

The analysis proposed in this paper is based on the hypothesis that word order variation in three-verb clusters in the Dutch language area does not involve parametrisation. As far as verb clusters are concerned, the dialects in this language area share the same grammar and this grammar allows for a certain amount of variation. Generative theory should explain why the order 2–1–3 is impossible and why the wellformedness of 2–3–1 and 3–1–2 depends on the type and hierarchy of the verbs in the cluster. On this view, the grammatical orders are, in principle, freely available for all speakers with this grammar, the actual choice in judgement tasks and language use being determined by sociolinguistic factors. These may well include geographical and social norms as

well as considerations of register and context.[3] For these reasons, I take the view that generative linguistics and sociolinguistics are complementary in that it is the task of sociolinguistics to describe and explain the patterns of variation that occur within a linguistic community given the theoretical limits of this variation uncovered by generative linguistics.

This paper is organised as follows. Section 2 provides an overview of the distribution of the various word orders in verb clusters in the Dutch language area. Section 3 compares some theoretical options to account for word order variation in verb clusters. Section 4 offers a new analysis that explains the generalisations described in Section 2 and shows how the variation space is determined by generative theory. Section 5 discusses the questions that fall outside the scope of generative theory and Section 6 concludes the paper.

2. Word order variation in three-verb clusters

2.1 Empirical findings

In the SAND-project (Syntactic Atlas of the Dutch Dialects) we have tested three different types of verb clusters in two different rounds.[4] As a pilot study, we sent out a questionnaire to 368 informants in 321 distinct locations with a reasonably even distribution over the Dutch language area. The three types of verb cluster are given in (2). For all logically possible orders of each type, we asked the informants to indicate whether this order occurred in their dialects.

(2) a. Modal – Modal – Verb (Mod-Mod-V)
   Ik weet dat Jan hard moet kunnen werken
   I   know that John hard must can.INF work.INF
   "I know that John should be able to work hard."

 b. Modal – Auxiliary – Verb (Mod-Aux-V)
   Jan   weet dat hij voor drie uur de wagen moet
   John knows that he   before three hour the car   must
   hebben   gemaakt
   have.INF repaired
   "John knows that he must have repaired the car before three o'clock."

 c. Auxiliary – Aspectual – Verb (Aux-Asp-V)
   Ik weet dat hij is gaan   zwemmen
   I   know that he   is go.INF swim.INF
   "I know that he went out to go swimming."

Table 1. Word orders in three-verb clusters in dialects of Dutch

Type of cluster	Order	Written questionnaire	Oral interviews	Geographic pattern
1Mod-2Mod-3V	1–2–3	273/321 = 85.0%	239/267 = 89.5%	yes
	1–3–2	35/321 = 9.7%	35/267 = 13.1%	yes
	2–1–3	1/321 = 0.3%	–	–
	2–3–1	4/321 = 1.2%	4/267 = 1.5%	no
	3–1–2	72/321 = 22.4%	81/267 = 30.0%	yes
	3–2–1	26/321 = 8.0%	37/267 = 13.9%	yes
1Mod-2Aux-3V	1–2–3	199/321 = 62.0%	79/267 = 29.6%	yes
	1–3–2	148/321 = 46.1%	146/267 = 54.7%	yes
	2–1–3	5/321 = 1.6%	–	–
	2–3–1	2/321 = 0.6%	–	–
	3–1–2	289/321 = 90.0%	174/267 = 65.2%	yes
	3–2–1	84/321 = 26.2%	48/267 = 18.0%	yes
1Aux-2Asp-3V	1–2–3	261/321 = 81.3%	158/267 = 59.2%	yes
	1–3–2	2/321 = 0.6%	–	–
	2–1–3	2/321 = 0.6%	–	–
	2–3–1	159/321 = 49.5%	109/267 = 40.8%	yes
	3–1–2	10/321 = 3.7%	3/267 = 1.1%	no
	3–2–1	40/321 = 12.5%	46/267 = 17.2%	yes
1Aux-2Mod-3V	1–2–3		180/267 = 67.4%	yes
	1–3–2		6/267 = 2.2%	yes
	2–1–3		0/267 = 0.0%	–
	2–3–1		26/267 = 9.7%	yes
	3–1–2		4 / 267 = 1.5%	no
	3–2–1		32 /267 = 12.0%	yes

We checked the results of the first round in 267 oral interviews with two and sometimes three informants at each location (cf. Cornips & Jongenburger 2001 and Cornips & Poletto 2005 for the methodology used in the SAND-project). In the oral interviews, we did not present all logically possible orders. The orders that were systematically absent in the results of the written questionnaire and that had never been reported in the dialect syntax literature were left out for practical reasons. The results of both studies are summarised in Table 1.[5]

2.2 Linguistic distribution

Table 1 compares the number of locations in which a certain cluster order is possible with the total number of locations involved in the study. A lower per-

centage for a particular order crucially does not mean that this order is less grammatical, it means that the order occurs in fewer dialects. In principle, an order that occurs only once should be considered grammatical.

However, for certain orders with very low percentages, the question must be asked whether they really occur. They could be ungrammatical or involve a different construction and the fact that their occurrence rate is not zero could be due to 'noise' in the data. I use the following criteria to evaluate such cases: (1) the occurrence rate of the relevant order is 1.5 percent or lower (i.e. 4 occurrences or less); (2) the order has not been observed in the literature; (3) the order does not occur in spontaneous speech; (4) the dialects in which the order is claimed to occur do not cluster geographically. The more criteria a particular order meets, the more likely it is that this order does not exist for a particular cluster type.

The fourth criterion, geographic clustering, is hard to apply without advanced statistical cluster analysis. I hope to carry out such an analysis in the near future. It is clear, however, that dialects in the Dutch language area that have a syntactic variant in common show a strong tendency to cluster geographically and that isolated occurrences are very rare (cf. Barbiers et al. 2005 for many examples and maps of such clusterings). For the purposes of this paper I will use the notion of geographic clustering in a more intuitive way, by indicating which geographic area a cluster of dialects share, e.g. along a border of a river or a lake, within a province etc. In combination with the other three criteria, then, lack of geographic patterning can be taken as an indication for the non-existence of a certain order.

On the basis of Table 1, the following observations can be made with respect to the distribution of word orders over the three different types of cluster. The geographic distribution of these word orders is discussed in Section 2.3

1–2–3. This order is possible and common in all three types of cluster.

1–3–2. This order is quite common in the type Mod-Aux-V, occurring in about 50% of the locations. In the type Mod-Mod-V it is much less common but still occurs in about 10% of the locations, which show a clear geographic pattern. In the type Aux-Asp-V, the 1–3–2 order is almost absent in the results of the postal survey. Therefore, it was not tested in the oral interviews. Presumably, this is a mistake, for the following reasons. In the oral interviews the informants were asked to translate some sentences of the type given in (3).

(3) Vertel maar niet wie zij had kunnen roepen
     Tell   just   not who she had can.INF call.INF
     "Just don't say who she could have called."

The cluster in this sentence contains a modal instead of an aspectual auxiliary as the second highest verb, but the results in Table 1 (third and fourth cluster type) strongly suggest that this difference does not influence the word order variation in the cluster (cf. also Section 4 for discussion). The sentences to be translated were presented in the Standard Dutch 1–2–3 order. In six locations (2.2%), the informants spontaneously changed the order into 1–3–2, which may indicate that this is a possible order in their grammars. It can also be claimed that these six locations cluster geographically, as they are all located around the former Zuiderzee (cf. Map 5).[6] I therefore tentatively conclude that the order 1–3–2 is possible in clusters of the type Aux-Mod-V and hence also in clusters of the type Aux-Asp-V.

2–1–3. This order is almost categorically absent in all three types. It does not show up in spontaneous speech or translations either. As this is true for verb clusters outside the Dutch language area as well, it is likely that we are dealing here with a general linguistic restriction for which syntactic theory should provide an explanation. This result confirms existing claims in the literature (Bloemhof 1979; Zwart 1995; Broekhuis 1997).[7]

2–3–1. This order is quite common in the type Aux-Asp/Mod-V, and almost absent with the other types. The four cases of 2–3–1 in clusters of the type Mod-Mod-V (1.5%) do not show a geographic pattern. They are isolated occurrences located in four non-adjacent provinces, Noord-Holland, Noord-Brabant, Groningen and Overijssel. They neither show up in spontaneous speech nor in the literature. It is therefore reasonable to conclude that 2–3–1 can only occur if there is a perfective auxiliary and this perfective auxiliary is the highest verb in the cluster, as was already suggested in Zwart (1995) (cf. also Den Dikken 1994; Broekhuis 1997). Again, we seem to be dealing with a linguistic factor that syntactic theory should explain.

3–1–2. This order is quite common in clusters of the type Mod-Mod-V and Mod-Aux-V. It is very rare in the type Aux-Asp-V (3 occurrences in the oral interviews, i.e. less than 1.5%) and in the type Aux-Mod-V (4 occurrences in the oral interviews = 1.5%). Although the attested cases occur along the eastern border of the language area (cf. Maps 4 and 5), they are far from each other and relatively isolated. Because of the low occurrence rate and the absence of a clear geographic pattern, I assume that this order is not possible for the type Aux-Asp/Mod-V.[8] This is the third categorical restriction that should be explained.[9]

3–2–1. This order is possible with all three types of cluster.

The word order possibilities for each cluster type are summarized in Table 2. Syntactic theory should explain why 2–1–3 is categorically impossible and

Table 2.

Type of cluster	Order	
1Mod-2Mod-3V	✓	1–2–3
	✓	1–3–2
	*	2–1–3
	*	2–3–1
	✓	3–1–2
	✓	3–2–1
1Mod-2Aux-3V	✓	1–2–3
	✓	1–3–2
	*	2–1–3
	*	2–3–1
	✓	3–1–2
	✓	3–2–1
1Aux-2Asp/Mod-3V	✓	1–2–3
	✓	1–3–2
	*	2–1–3
	✓	2–3–1
	*	3–1–2
	✓	3–2–1

why clusters of the type Aux-Asp/Mod-V differ from the other two in allowing 2–3–1 and not allowing 3–1–2.

2.3 Geographic distribution

For ease of reference, Map 1 provides an overview of the language area including the names of the provinces referred to in the text and the names of all the locations involved in the SAND project.

2.3.1 *Cluster type Modal-Modal-Verb*
Mod-Mod-V. Map 2 provides the results of the oral interviews (Table 1, column 4; *moet kunnen werken* 'must can.ɪɴꜰ work.ɪɴꜰ'). The various orders were presented to the informants in their local dialects. The informants were asked whether these sentences occurred in their dialects. Map 2 shows that there are two primary orders: 1–2–3 and 3–2–1. The 1–2–3 order (symbol: △) is the most common order in the entire language area and the only option in many dialects, in particular in Belgium. The 3–2–1 order (symbol: □) occurs in the northern provinces (Friesland, Groningen, Drenthe and Noord-Holland) and

Map 1. Language area of the Dutch dialects*

* The Dutch speaking part of Belgium includes the following provinces (from west to east): West- and Oost-Vlaanderen, Vlaams Brabant and Antwerpen, and Vlaams Limburg. The area to the south west of West-Vlaanderen is called Frans-Vlaanderen. It belongs to France. Very old speakers in this area still speak Dutch dialects. The Netherlands includes the following provinces: Friesland, Groningen, Drenthe, Overijssel, Gelderland, Utrecht, Noord-Holland, Zuid-Holland, Zeeland, Noord-Brabant, Limburg. There are no data from the province of Flevoland (to the north of Utrecht and Gelderland), as it was impoldered recently and does not have inhabitants that fit our criteria for informants.

☐ 3V - 2Modal - 1Modal
    37 locations
△ 1Mod - 2Mod - 3V
    239 locations
❘ 3V - 1Modal - 2Modal
    81 locations
─ 1Modal - 3V - 2Modal
    35 locations
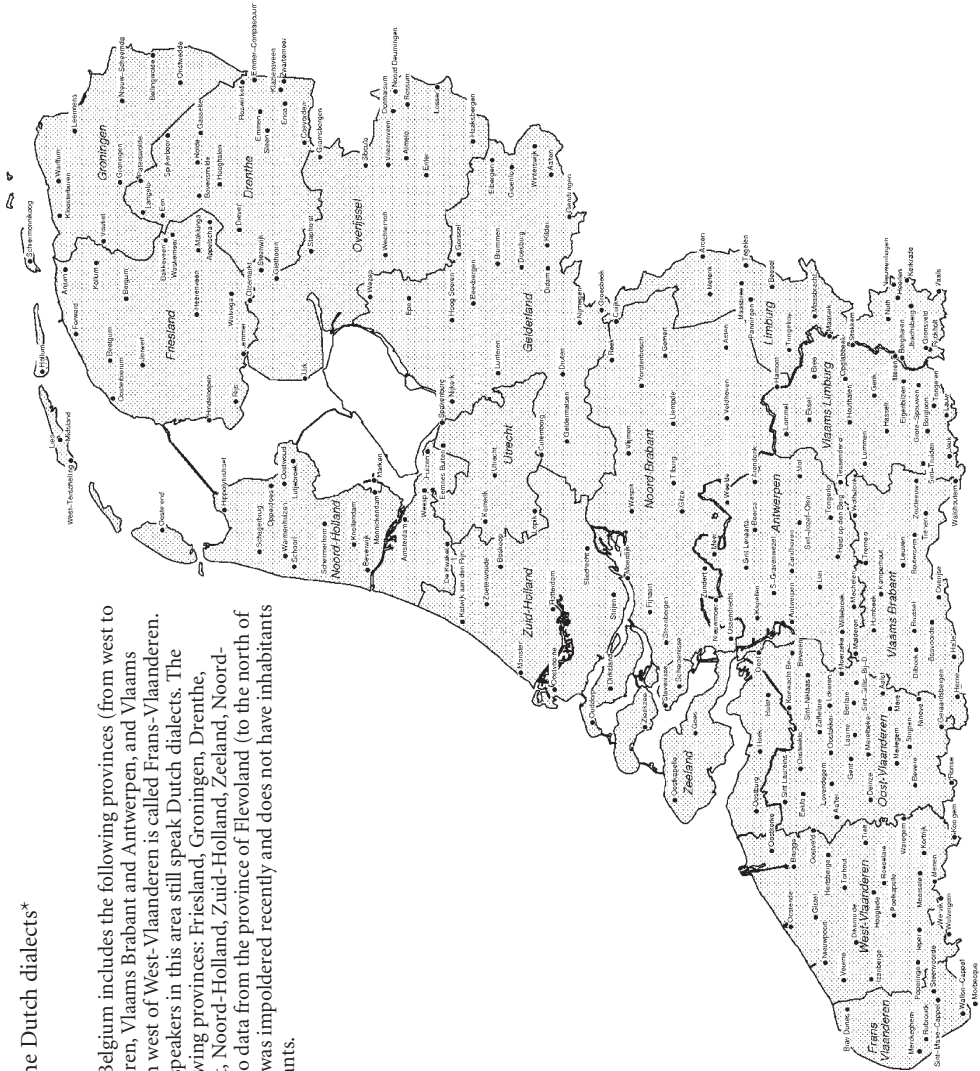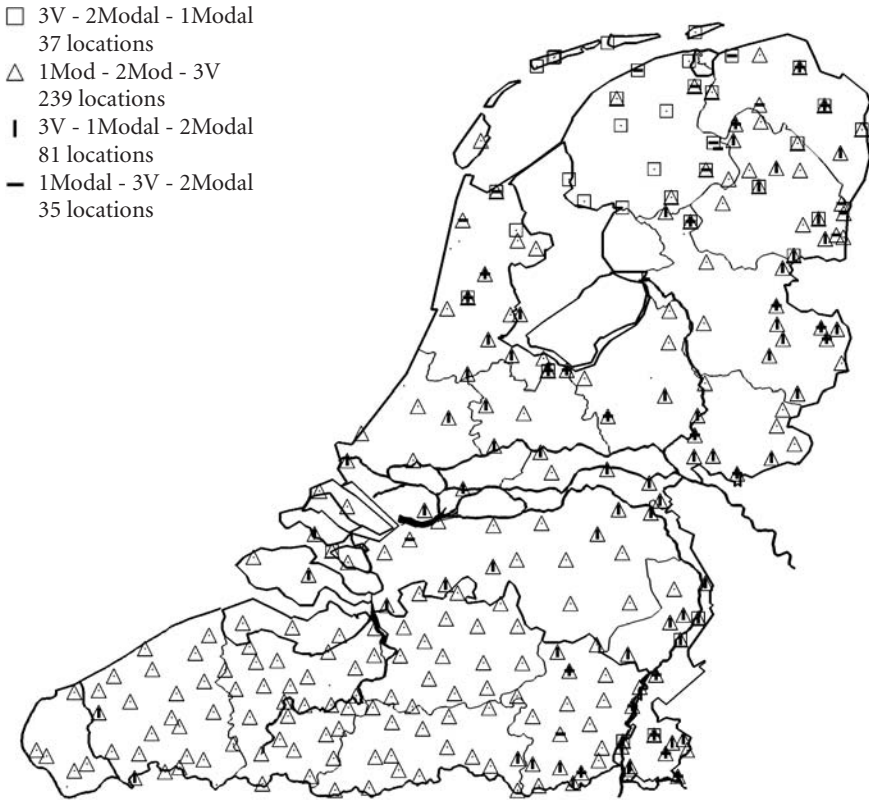
Map 2. Distribution of word orders in clusters of the type Modal-Modal-Verb

in Dutch Limburg. In Friesland, it is often the only order that is claimed to oc-
cur, in the other provinces there are several alternatives. The orders 3–1–2 and
1–3–2 never occur as the only option in a dialect and can therefore be consid-
ered secondary orders. The 3–1–2 (symbol: |) is very rare in Belgium, occurring
a couple of times in Vlaams Limburg and twice in West-Vlaanderen.[10] 3–1–2 is
quite common in the Netherlands, in particular in but certainly not restricted
to the eastern zone. The 1–3–2 order (symbol: —) has a distribution quite sim-
ilar to 3–1–2, but a much lower occurrence rate, and it hardly occurs at all in
the western part of the Netherlands, as opposed to the 3–1–2 order.

2.3.2 *Cluster type Modal-Auxiliary-Verb*
Mod-Aux-V. Map 3 provides the results of the oral interviews (Table 1, column
4; *moet hebben gemaakt* 'must have.ɪɴꜰ made.ᴘᴀʀᴛ'). The various orders were

□ 3V - 2Aux - 1Modal
48 locations
△ 1Modal - 2Aux - 3V
79 locations
| 3V - 1Modal - 2Aux
174 locations
— 1Modal - 3V - 2Aux
146 locations
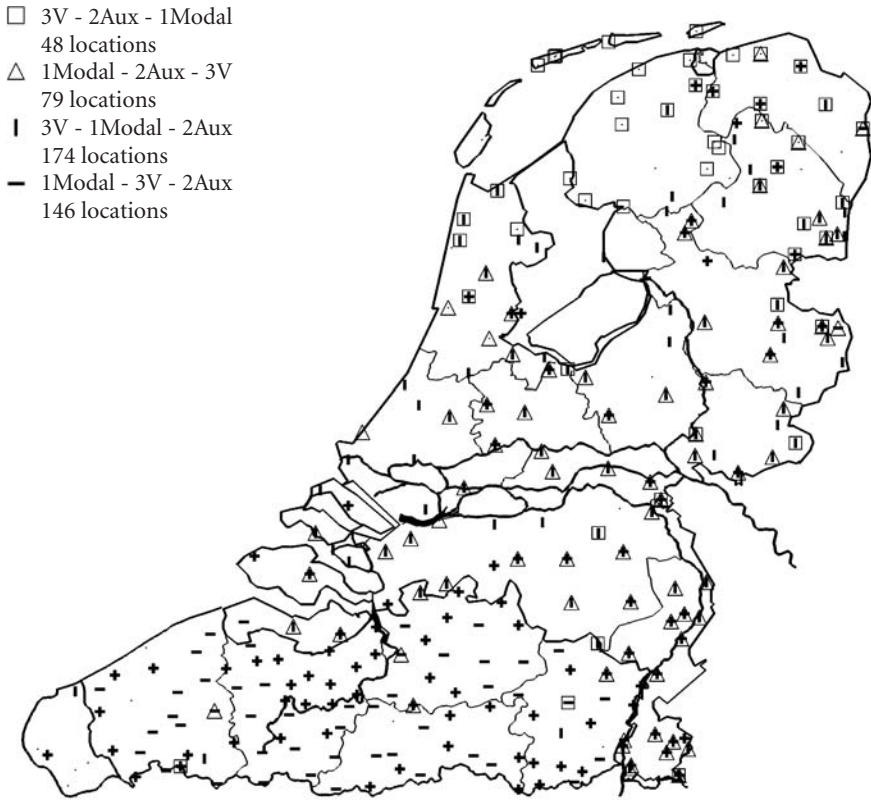
Map 3. Distribution of word orders in clusters of type Modal-Auxiliary-Verb

presented to the informants in their local dialects. The informants were asked whether these sentences occurred in their dialects. There are some striking differences with the word order variation in the cluster type Mod-Mod-V. The two secondary orders of that cluster type, 3–1–2 and 1–3–2, are the primary orders here in that they have the highest number of occurrences and that they are the only option in quite a few dialects. The 1–3–2 order (symbol: —) is very dominant in Belgium. Many dialects in Oost-Vlaanderen and Vlaams Limburg allow the 3–1–2 order (symbol: |) in addition to 1–3–2, but 3–1–2 rarely is the only option in Belgium.[11] The 1–2–3 order (△) is much less common than in the type Mod-Mod-V. In particular, 1–2–3 is almost absent in Belgium. In the Netherlands, 1–2–3 occurs in all provinces except Friesland, but seldom as the only possible order. The 3–2–1 order (□) again is mainly found in the northern

Map 4. Distribution of word orders in clusters of the type Auxiliary-Aspectual-Verb

and eastern dialects. In Friesland, this order is often the only option. Locations with this order outside of Friesland usually have one or more alternative orders.

2.3.3 *Cluster type Auxiliary-Aspectual-Verb*

Aux-Asp-V. Map 4 provides the results of the oral interviews (Table 1, column 4; *is gaan zwemmen* 'is go.inf swim.inf'). The various orders were presented to the informants in their local dialects. The informants were asked whether these sentences occurred in their dialects. The most striking feature of this distribution is the large numbers of dialects that have the order 2–3–1 (symbol: ○), an order that is entirely absent in the two-cluster types discussed above. This order occurs both in Belgium and the Netherlands. In Belgium, 2–3–1 is the only option in the majority of dialects. In the periphery (Vlaams Limburg and western West-Vlaanderen) we find a number of dialects that allow 1–2–3 in ad-

dition to 2–3–1. In the Netherlands, 2–3–1 primarily occurs in the western part of the country (Zeeland, western Noord-Brabant, Zuid- and Noord-Holland), but there are also some attestations in the central and eastern part. Almost all dialects with 2–3–1 in the Netherlands have 1–2–3 as an alternative option. The 1–2–3 order (△) rarely occurs as the only possibility in the Belgian dialects. In contrast with this, the majority of dialects in the Netherlands have 1–2–3 as the only possible order. The 3–2–1 order (□) is mainly restricted to dialects of the northern provinces, including Friesland, Groningen and Noord-Holland.[12] In some of these dialects, this is the only option. The order 1–3–2 was not tested in the oral interviews (cf. Section 2.2 for motivation). However, in the description of Map 5 I conclude that 1–3–2 is a possible order for this cluster type. Finally, the order 3–1–2 (|) occurs in only three dialects, located along the eastern border with Germany. In Section 2.2, I concluded that this order does not exist for this cluster type, because of the low occurrence rate and the absence of a clear geographic pattern.

2.3.4 *Cluster type Auxiliary-Modal-Verb*

Aux-Mod-V. Map 5 provides the results of a translation task (Table 1, column 4, fourth cluster type). The informants were asked to translate the sentence (*Vertel maar niet wie zij*) *had kunnen roepen* lit. 'tell just not who she' had can.ɪɴꜰ call.ɪɴꜰ 'You better don't say who she could have called.' They were not asked to give alternative orderings. As a result, most of the locations have only one reported order here. This map has much in common with Map 4, both with respect to the different orders that occur and to their distribution. It is clear that there are three main orders. 3–2–1 only occurs in the north (Friesland, Groningen and Drenthe). 1–2–3 occurs in all provinces of the Netherlands except Friesland and Groningen. 1–2–3 also occurs in Belgium, in particular in the provinces of West-Vlaanderen, Antwerpen and Vlaams Limburg. 2–3–1 mainly occurs in Oost-Vlaanderen and Vlaams-Brabant; in some Antwerp dialects it is an alternative for 1–2–3. As was already noted above, the order 1–3–2 shows up in this translation test. Since the informants spontaneously changed the order that was offered and since the locations where this happened show a geographic pattern (in an area along the former Zuiderzee), it is reasonable to assume that the occurrence of this order is real. Finally, the order 3–1–2 is again extremely rare (4 locations = 1.5%). Three of them occur along the eastern border of the Netherlands, one of them on the northern island of Terschelling, so there is no clear geographic pattern (cf. Note 7). I therefore assume that the order 3–1–2 does not exist for this type of cluster.

◻ 3V - 2Mod - 1Aux
32 locations
△ 1Aux - 2Mod - 3V
180 locations
○ 2Mod - 3V - 1Aux
26 locations
❙ 3V - 1Aux - 2Mod
4 locations
— 1Aux - 3V - 2Mod
6 locations

Map 5. Distribution of word orders in clusters of the type Auxiliary-Modal-V

The many similarities between Map 4 and Map 5 suggest that there is no principled difference between cluster type Aux-Mod-V and cluster type Aux-Asp-V. One difference between Map 4 and Map 5 requires further research. The distribution of 2–3–1 is much more restricted on Map 5 than on Map 4, where it extends to the north western part of the Netherlands. If I am correct in assuming that the two cluster types allow the same orders, these different geographic distributions must be due to the fact that one is the result of a judgement task (in disguise) and the other of a translation task. A tentative explanation of this difference could be that informants tend to stick to the presented word order in translation tasks, even if that is not the preferred order in their dialects. This would imply that the wider distribution of 2–3–1 on Map 4 is the most reliable.

3. Some theoretical options

The purpose of this section is to evaluate some possible theoretical analyses of three-verb clusters to establish whether they adequately generate all and only the attested orders discussed in the previous section.[13] I start with the simplest analysis possible, one in which word order variation in verb clusters is the result of base-generating different but equivalent syntactic structures. Two structures are equivalent when the embedding/containment relations are identical. It is natural to assume that alternative structures should be syntactically equivalent, since only syntactically equivalent structures can receive identical interpretations at Logical Form. The structures in (4) exhaust the various options.

(4) a. 1–2–3

```
        VP1
       /   \
      V1    VP2
      |    /   \
      |   V2    VP3
      |   |     |
      |   |     V3
      1   2     3
```

b. 1–3–2

```
        VP1
       /   \
      V1    VP2
      |    /   \
      |  VP3    V2
      |   |     |
      |   V3    |
      1   3     2
```

c. 2–3–1

```
          VP1
         /   \
       VP2    V1
      /   \    |
     V2   VP3  |
     |    |    |
     |    V3   |
     2    3    1
```
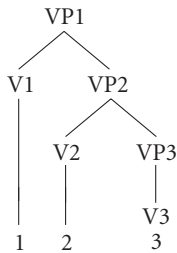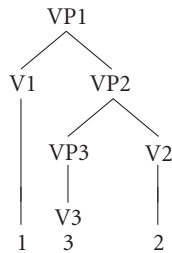
d. 3–2–1
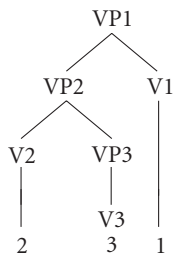
```
          VP1
         /   \
       VP2    V1
      /   \    |
    VP3    V2  |
     |     |   |
     V3    |   |
     3     2   1
```
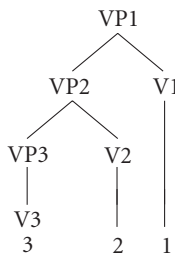
The base generation analysis rules out the orders 2–1–3 and 3–1–2. It is impossible to draw a tree such that VP1 contains VP2 and VP2 contains VP3 and the order is 2–1–3 or 3–1–2. Note that optionality of the type depicted in (4) does not pose a problem for the minimalist framework, as these structures do not involve movement. A central hypothesis of minimalism is that language design is economical and that the number of operations is kept to a minimum, i.e. operations only apply when necessary. If a movement operation is optional, it is not necessary and hence it should not occur. All structures in (4) are the

result of two 'Merge' operations: merge V2 and V3, and then merge V1 with V2 and V3. Therefore, these structures are equally costly, and less costly than movement analyses of three-verb clusters. In the absence of direct evidence for movement in verb clusters, this seems to be an attractive analysis. Unfortunately, it must be rejected given the empirical results of the previous section, as it wrongly rules out the quite popular 3–1–2 order. This order thus requires a 'Movement' analysis.

The word order variation in (4) derives from the assumption that a syntactic projection can be either head-first or head-final: a head can either precede or follow the projection with which it is merged. The structure in (4a) is uniformly head-first, the structure in (4d) is uniformly head-final, and the structures in (4b, c) are mixed. A central assumption of the antisymmetry framework (Kayne 1994) is that all structures are head-first underlyingly. Even if the structures in (4) were possible, they would all yield the same linear order 1–2–3, because linear order derives from asymmetric c-command in the antisymmetry framework.[14] In all structures V1 asymmetrically c-commands V2 and V2 asymmetrically c-commands V3. All orders except 1–2–3 should therefore be derived by optional VP movement. Like any other movement operation in the antisymmetry framework, this movement is leftward. A VP moves up to a position immediately preceding the next verb higher up. Let us assume that this position is the specifier of that higher verb:[15,16]

(5)
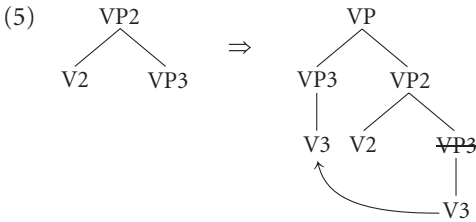
The optional leftward VP movement analysis automatically derives the five orders found in verb clusters, and correctly rules out the 2–1–3 order. The various derivations are given in (6).

(6) a. No movement   Order
         [$_{VP1}$ V1 [$_{VP2}$ V2 [$_{VP3}$ V3]]]   1–2–3
     b. VP3 to SpecVP2
         [$_{VP1}$ V1 [$_{VP2}$ [$_{VP3}$ V3] V2 ~~[$_{VP3}$ V3]~~]]   1–3–2
     c. VP2 to SpecVP1
         [$_{VP1}$ [$_{VP2}$ V2 [$_{VP3}$ V3]] V1 ~~[$_{VP2}$ V2 [$_{VP3}$ V3]]~~]   2–3–1

    d.   VP3 to SpecVP2, VP2 to SpecVP1

        [$_{VP1}$ [$_{VP2}$ [$_{VP3}$ V3] V2 [~~$_{VP3}$ V3~~]]] V1 [~~$_{VP2}$ [$_{VP3}$ V3] V2 [$_{VP3}$ V3]~~] 3–2–1

    e.   VP3 to SpecVP2, VP3 to SpecVP1

        [$_{VP1}$ [$_{VP3}$ V3] V1 [$_{VP2}$ [~~$_{VP3}$ V3~~] V2 [~~$_{VP3}$ V3~~]]]   3–1–2

The 2–1–3 order would require movement of VP2 to SpecVP1, leaving VP3 behind. This is impossible, because VP2 contains VP3 and carries VP3 along with it when it moves. Put differently, the base generation analysis and the movement analysis are equivalent as far as ruling out 2–1–3 is concerned. In neither of these analyses can this order be generated without changing the original embedding relations. The optional leftward VP movement analysis is superior to the base generation analysis in that it correctly rules in the 3–1–2 order.[17]

    Notice that a head movement analysis would not be a good alternative. First, if movement in V-clusters were head movement, 2–1–3 could be derived by moving V2 to the left of V1. Secondly, the head movement analysis makes a false prediction for particle stranding, as Koopman and Szabolcsi (2000) show. Particle stranding (more generally, argument and adjunct stranding) is impossible in verb clusters (7a), whereas such stranding is possible in cases of real head movement such as V2 (7b).

(7) a. *dat Jan   de koek eten   mag op ~~eten~~

           that John the cake eat.INF may up eat.INF

           "That John is allowed to eat up the cake."

      b. Jan   eet de koek op ~~eet~~

           John eats the cake up eats

           "John is eating up the cake."

The optional leftward VP movement analysis yields good results, but there are two further issues that one would want to clarify. Firstly, only clusters of the type Aux-Mod/Asp-V allow the order 2–3–1 and disallow 3–1–2. Secondly, we want to understand the conditions under which movement in verb clusters is possible. These two interrelated issues are discussed in the next section.

4. Analysis

The optional leftward VP movement analysis correctly generates all and only the five orders that occur in the Dutch language area. Our next task is to understand the conditions on VP movement and to explain why only clusters of the type Aux-Mod/Asp-V allow the order 2–3–1 and disallow 3–1–2. I argue that

VP movement, like any other type of movement, requires agreement between the attractor, in this case Aux, Mod or Asp, and the constituent that moves, in this case VP.[18] Modal and aspectual auxiliaries agree with the dependent main verb but not with dependent modal, aspectual or perfective auxiliaries. Perfective auxiliaries may agree both with the dependent main verb and with dependent auxiliaries, but only when they are perfective.

In the minimalist program, movement is considered to be an imperfection of the language system and tied to another imperfection, uninterpretable morphosyntactic features (uF) (cf. Chomsky 2001; Pesetsky & Torrego 2001), such as agreement features on verbs and adjectives. When a derivation reaches the level of semantic interpretation, Logical Form (LF), there should not be any uninterpretable features left, or the derivation will crash, giving rise to an ungrammatical sentence. Uninterpretable features are deleted under a relation of *Agree*. An uninterpretable feature uF looks into its c-command domain to find the closest agreeing interpretable feature iF. If it finds one, it is deleted. Agree is a necessary condition on movement, i.e. without Agree there is no movement. Uninterpretable features trigger movement of an element with an agreeing interpretable feature. In many cases, such movement is optional.

If agreement is a precondition for movement, then VP movement inside verb clusters should involve agreement as well. The question is which features are relevant for agreement here. A main verb denotes an event, a modal or aspectual auxiliary does not, but needs to combine with an expression denoting an event. I therefore assume that main verbs have a feature 'iEvent', whereas modal and aspectual auxiliaries have a feature 'uEvent'.[19] Let us further assume that perfective auxiliaries have a feature 'uPerfective', while perfect participles have a feature 'iPerfective'. The idea behind this assumption is that perfect participles denote perfectivity, whereas perfective auxiliaries do not. Perfective auxiliaries need to combine with an expression denoting perfectivity. Independent evidence for the latter assumption will be provided below.

For clusters of the type 1Mod-2Mod-3V, I propose the analysis in (8). As indicated, in every step of the derivation there is an agreement relation between the main verb and one of the auxiliaries. The iEvent feature of the main verb checks the uEvent features of the auxiliary verbs. The agreement relation between uEvent and iEvent makes movement possible, but this movement is optional.

(8) Derivations 1Mod-2Mod-3V: moet   kunnen   zwemmen
                                must   can.ɪɴғ   swim.ɪɴғ

Step 1: Merge *kunnen* and *zwemmen*
[$_{VP2}$ 2-kunnen [$_{VP3}$ 3-zwemmen]]
        uEvent   iEvent
        └─────────────────┘ Agree → Deletion of uEvent of *kunnen*

Option 1: no movement   → order 2–3
Option 2: move VP3 to SpecVP2 → order 3–2

Step 2: Merge *moet*
Continuation option 1:
[$_{VP1}$ 1-moet [$_{VP2}$ 2-kunnen [$_{VP3}$ 3-zwemmen]]]
        uEvent   ~~uEvent~~   iEvent
        └──────────────────────────────┘ Agree → Deletion of uEvent of *moet*

Option 1a: no movement   → Final order 1–2–3
Option 1b: move VP3 to SpecVP1 → Final order 3–1–2

Continuation option 2:
[$_{VP1}$ 1-moet [$_{VP2}$ [$_{VP3}$ 3-zwemmen] 2-kunnen [$_{VP3}$ ~~3-zwemmen~~]]]
        uEvent   iEvent   ~~uEvent~~
        └──────────────────────┘ Agree → Delete uEvent of *moet*

Option 2a: no movement   → Final order 1–3–2
Option 2b: move VP3 to SpecVP1, pied-piping VP2 → Final order 3–2–1
Option 2c: move VP3 to SpecVP1, no pied-piping   → Final order 3–1–2

Option 2b is a case of pied-piping. We are dealing with pied-piping when more
is moved than strictly necessary. In Option 2b, there is an agreement rela-
tion between VP3 and V1, thus VP3 can be moved. When it moves it may
carry along VP2. Thus, two distinct types of optionality cause variation here:
(i) optionality of movement; (ii) optionality of pied-piping (cf. Koopman &
Szabolcsi 2000).

Both types of optionality are needed independently in the grammar of
Dutch, they are not construction-specific stipulations. Another example of op-
tionality of movement is that constituents with focus accent can, but need, not
be preposed.

(9) a. We hebben gisteren ᴊᴀɴ   gebeld
         We have   yesterday ᴊᴏʜɴ called
         "Yesterday, we called ᴊᴏʜɴ."
     b. ᴊᴀɴ   hebben we gisteren   gebeld
         ᴊᴏʜɴ   have   we yesterday called
         "ᴊᴏʜɴ we called yesterday."

Another example of optional pied-piping is given in (10). The DP contains a Wh-element (*wat* 'what') and Wh-elements must be preposed in Dutch in the unmarked case. This requirement can either be met by moving the entire DP (10a; pied-piping), or by moving just *wat* 'what' from SpecDP (10b; no pied-piping). The first option (10a) is parallel to Option 2b in (8), where movement of VP3 causes movement of the entire VP2 containing VP3. The second option (10b) is parallel to Option 2c in (8), where VP3 is moved from SpecVP2 and VP2 remains in situ.

(10)   a.   [$_{DP}$ Wat [voor een [$_{NP}$ fiets]]] heb je   [$_{DP}$ ~~wat~~   [~~voor~~
             What   for   a   bike   have you   what   for
             ~~een~~ [$_{NP}$ ~~fiets~~]]] gekocht
             a   bike   bought
             "What kind of bike have you bought?"

       b.   [Wat] heb je   [$_{DP}$ ~~wat~~ [voor een [$_{NP}$ fiets]]] gekocht?
             What have you   what   for   a   bike   bought
             "What kind of bike have you bought?"

The analysis proposed in (8) rules out the unattested orders 2–1–3 and 2–3–1 automatically. The crucial point is that V1 and V2 do not agree, as they both have an uninterpretable feature. Therefore, V1 does not attract VP2.[20] As we have seen, the order 2–1–3 would be out even if there were agreement, since movement of VP2 will always carry along VP3, giving rise to either the 2–3–1 or the 3–2–1 order.

The analysis for clusters of the type 1Mod-2Aux-3V is essentially the same. The presence of the Perfective features is irrelevant here as it does not change the Agreement relations.

(11)   Derivations 1Mod-2Aux-3V: moet hebben   gemaakt
                                  must have.INF made.PART
       [$_{VP1}$ 1-moet [$_{VP2}$ 2-hebben [$_{VP3}$ 3-gemaakt]]]
             uEvent   ~~uEvent~~   iEvent
                              ~~uPerfective~~   iPerfective
                              |_____| Agree in Step 1
             |_____| Agree in Step 2

The presence of the Perfective features does make a difference, however, when the hierarchy is 1Aux-2Asp-3V or 1Aux-2Mod-3V, as we will see shortly. I treat these two clusters as one type, since they both have 1–2–3 as the main order, they both allow the 2–3–1 order, unlike the Mod-Mod-V and the Mod-Aux-V type (cf. Table 1). Moreover, they both allow the 3–2–1 order in a significant

number of cases. I assume that they both allow the order 1–3–2 too, but not 3–1–2 (cf. the discussion in Sections 2.2 and 3). Finally, the geographic distribution of the various orders is very similar for both types of cluster (cf. Section 2.3.3 and 2.3.4).

I assume that the feature specification of the verbs involved is as in (12).

(12) Feature specification Aux-Mod/Asp-V
      [VP1 1-heeft [VP2 2-kunnen [VP3 3-roepen]]]
           has   can.INF   call.INF
      [VP1 1-is [VP2 2-gaan [VP3 3-zwemmen]]]
           is   go.INF   swim.INF
           | | |
           uPerfective uEvent   iEvent
                        iPerfective

This feature specification requires independent justification. Firstly, the claim that Perfective is uninterpretable on perfective auxiliaries and interpretable on participles is supported by the fact that attributively used participles have a perfective interpretation even when a perfective auxiliary is lacking (13). A perfective auxiliary, on the other hand, cannot express perfectivity in the absence of a perfect participle, but together with a perfect participle it can do so.

(13) de geroepen   ober
      the called-PART   waiter

Secondly, the modal *kunnen* 'can.INF' and the aspectual gaan 'go.INF' in (12) are perfective despite their misleading infinitival form. In fact, it is surprising that we have an infinitival form here, as perfective auxiliaries usually require a participle as their complement rather than an infinitive (14).

(14) Jan   heeft dat   nooit gekund   / *kunnen
      John has   that never could.PART   can.INF
      "John has never been able to do that."

It is a peculiar property of Standard Dutch and many of its dialects that the presence of a third verb in constructions such as (14) requires the modal to have an infinitival form. In north eastern dialects of Dutch this so called *Infinitivus pro Participio* effect (IPP) can be absent, and there we see that the modal is perfective indeed (15). Moreover, the interpretation of the modal in (15) is perfective too.

(15)   Jan   had   het   hele   brood   wel   op-eten   wild.
       John   had   the   whole   bread   AFFIR   up-eat   wanted.PART
       "John would have liked to eat up the entire bread."

The proposed feature specification yields the following derivations and orders:

(16)   Derivations 1Aux-2Mod/Asp-3V: heeft   kunnen   roepen
                                     heeft   can.INF   call.INF

      Step 1: Merge *kunnen* and *roepen*
      [$_{VP2}$ 2-kunnen [$_{VP3}$ 3-roepen]]
           iPerf
           uEvent   iEvent
           |_____| Agree → Delete uEvent

      Option 1: No movement   → Order 2–3
      Option 2: VP3 moves to SpecVP2 → Order 3–2

      Step 2: Merge *heeft*
      Continuation option 1:
      [$_{VP1}$ 1-heeft [$_{VP2}$ 2-kunnen [$_{VP3}$ 3-roepen]]]
                  ~~uEvent~~   iEvent
           uPerf   iPerf
           |_____| Agree → Delete uPerf of *heeft*

      Option 1a: No movement   → Final order 1–2–3
      Option 1b: VP2 moves to SpecVP1, carrying along VP3 → Final order 2–3–1

      Continuation option 2:
      [$_{VP1}$ 1-heeft [$_{VP2}$ [$_{VP3}$ 3-roepen] 2-kunnen [$_{VP3}$ ~~3-roepen~~]]]
                  iEvent   ~~uEvent~~
           uPerf   iPerfective
           |_____| Agree → Delete uPerf of *heeft*

      Option 2a: No movement   → Final order 1–3–2
      Option 2b: VP2 moves to SpecVP1, pied-piping VP3 → Final order 3–2–1

A crucial difference with the other clusters is that the hierarchy 1Aux-2Asp/Mod-3V is the only one in which 1 and 2 agree, making [2–3] movement to SpecVP1 possible, yielding the order 2–3–1, which is impossible with the other hierarchies. The order 2–1–3 is still ruled out correctly. Since we are dealing with phrasal movement, movement of 2 obligatorily forces pied-piping of 3. Finally, the order 3–1–2 is correctly ruled out because the main verb (3) does not agree with the perfective auxiliary (1). Therefore the perfective auxiliary cannot attract the main verb.

The generative analysis proposed in this section explains the categorical properties of the distribution of word orders in verb clusters in the dialects of the Dutch language area, i.e. the properties that hold for all dialects involved, in particular, the fact that 2–1–3 does not occur, the fact that only 1Aux-2Asp/Mod-3V allows 2–3–1 and disallows 3–1–2 orders, and the fact that the remaining orders are indeed attested in the various cluster types. These facts follow from independently motivated principles. Constituent integrity (or X-bar theory) forces a VP that moves to carry along everything that is contained in it, making it impossible to move VP2 to SpecVP1 without VP3. Agreement as a general condition on movement operations rules out movement of VP2 (containing VP3) to SpecVP1 when there is no agreement between V1 and V2, and movement of VP3 to SpecVP1 when there is no agreement between V1 and V3. Thus, the proposed analysis adequately delimits the variation space.

In this account, the impossibility of the order 2–3–1 and the possibility of the order 3–1–2 in clusters of the type 1Mod-2Mod-3V and 1Mod-2Aux-3V are effectively reduced to lexical properties, in the spirit of the Minimalist Program. The lack of agreement between 1 and 2 or 1 and 3 in such clusters derives from the different morphosyntactic feature specifications of perfective auxiliaries on the one hand and modal and aspectual auxiliaries on the other.

When the conditions of constituent integrity and agreement are met, VPs can freely move through a cluster, but they need not do so. The word order variation attested in verb clusters across dialects thus derives from the optionality of the movement operations involved. This optionality is not reduced to lexical morphosyntactic features, but taken to be an inherent property of the grammatical system. This goes against the spirit of the Minimalist Program that all syntactic variation can be reduced to the lexicon (i.e. to morphosyntactic feature specification) or phonological form (the level of representation that is the input for the articulatory-perceptual interface). However, the accounts of optionality that have been proposed in the minimalist framework so far are not very convincing. In older versions of the framework, strong features triggered overt movement, whereas weak features did not. Optional movement occurred when a feature was optionally strong (cf. Henry 1995). I don't see how this is different from saying that movement is optional in certain cases. Similarly, in more recent versions of the Minimalist Program (Chomsky 2001), optionality of movement is derived from optional assignment of a so-called EPP-feature to an agreeing attractor.[21] Since the optionality is now in the assignment of the EPP-feature, these versions of the Minimalist framework also assume optionality as a property of the grammatical system. Thus, Minimalist

accounts of optional movement so far have failed to eliminate optionality from the grammar.

Rather than trying to eliminate optionality from the grammar, I propose that optionality of movement is an inherent property of the grammatical system. Movement does not have a trigger. It is freely available provided that it satisfies general conditions such as the ones discussed above. In a sense, then, this view on movement is a return to the Principle and Parameters (P&P) approach of movement (Chomsky 1981).

5.   Remaining questions about geographic and individual variation

If the hypothesis advanced here is correct, i.e. that all dialects investigated in the SAND-project have the same grammatical system for verb clusters, the individual and geographic variation attested must be due to extralinguistic factors.

Let us first consider individual variation. Maps 2–5 suggest that locations in the transitional zone between the Dutch and the German language area, i.e. the eastern part of the Netherlands, allow a larger number of word orders per cluster type than locations in other areas. For example, Map 3 shows that the majority of the dialects in the south-eastern part of the Netherlands (i.e. Dutch Limburg and the south-eastern part of North Brabant), have three orders: 1–2–3, 1–3–2 and 3–1–2. Most of the neighbouring Belgian dialects have only two orders, 1–3–2 and 3–1–2. Most of the neighbouring Hollandic dialects also allow only two orders, 1–2–3 and 3–1–2. If we were to assume that the Dutch Limburgian and North-Brabantish dialects involved have a grammar that allows three orders, whereas the neighbouring Belgian and Hollandic dialects have a grammar that allows only two orders, it would be a mere coincidence that the number of alternative orders is higher in the transitional zone.

An adequate analysis should therefore explain why speakers in the transitional zone show a greater freedom of word ordering in verb clusters. The following answer seems plausible. All orders that the grammatical system allows are, in principle, available for each speaker in the Dutch language area, but which orders a speaker actually uses or reports to occur in his dialect depends on the input from the environment. Speakers in transitional areas hear more different orders than speakers in non-transitional areas and hence allow a higher number of alternatives. In fact, this is a specific instantiation of the general generative hypothesis that universal grammar provides a large number of options from which a restricted set is selected during the process of language acquisition on the basis of actual input.

This explanation seems to be superior to one according to which the informants have reported the word orders that they are familiar with, not the word orders that they actually use in their own dialects. A problematic aspect of such an explanation is that it presupposes that informants are unable to distinguish between word orders that belong to their own dialect and word orders that belong to familiar but different dialects. Although this (in-)ability may vary per speaker and per syntactic construction and raises a number of questions (cf. Preston 2004), it is our experience in the SAND-project that informants often have quite strong, adequate and subtle intuitions about whether a certain construction that they know belongs to their own or to a familiar dialect. A typical response is: 'Yes, I know that they say that in village X five kilometres from here, but we don't say that.' Unfortunately, the explanation on the basis of familiarity with variants that occur in other dialects can not be tested here because there is not enough spontaneous speech for each dialect in the SAND data.

An explanation on the basis of familiarity is very similar to explanations on the basis of language contact or multilingualism. In fact, these three types of explanation can hardly be distinguished, as there are no objective linguistic tools to determine whether a certain order belongs to the same or a different grammar. When a native speaker judges that a certain construction does not occur in his dialect, this does not tell us whether this construction is excluded by the grammar of his dialect or merely unrealised. The same holds for the absence of a construction in a corpus. We may even doubt that there will ever be such tools. If all dialects are specific instantiations of one and the same grammar, then options that are missing in a certain dialect but not in others are merely unrealised, not impossible.

An observation that none of the above mentioned explanations account for is the fact that other parts of the transitional eastern zone of the Dutch language area are much less homogeneous than the south-eastern part. This shows that there is no simple relation between the type and number of orders that are reported to occur in a particular location and the type and number of orders that are reported to occur in the vicinity of that location. Consider, for example, the ten locations on Map 3 in the eastern part of Overijssel close to the German border. Three of these locations allow only one order, 3–1–2. Four locations allow two orders, of which two locations have 1–2–3 and 3–1–2, one location has 1–2–3 and 1–3–2, and one location has 3–1–2 and 3–2–1. Two of the ten locations allow three orders: 1–2–3, 1–3–2, 3–1–2. Finally, there is one location which has all four orders that are possible for this type. An illustration of the same problem is that there are also speakers in non-transitional zones that allow for three different orders. It is the task of sociolinguistics and socio-

dialectology to show whether, and to what extent, extra-linguistic factors such as social class, age, gender and geographic area play a role in this individual variation.

As for geographical variation, there are many clear patterns that require an explanation. I restrict myself to two observations. The first observation is that the Frisian area seems to be more homogeneous than the rest of the language area in that many Frisian locations only allow one order, 3–2–1, an order which is, in fact, relatively rare outside of Friesland. This can be captured by assuming that the other orders are grammatical but unrealised in the relevant dialects, possibly as a consequence of the standardisation of Frisian.[22] It could also indicate that these Frisian dialects share a yet to be identified linguistic property that the other dialects do not possess. A candidate property is obligatory pied-piping.

A second observation concerns the role that perfective auxiliaries play in the occurrence of certain orders. As a comparison of Maps 2 and 3 shows, the incidence of 1–3–2 orders increases considerably when the second highest verb is a perfective auxiliary instead of a modal or aspectual. Since the majority of these 1–3–2 orders is found in Belgium, a conceivable minimalist reflex could be to assume not just that the perfective auxiliary and the participle agree, but that the perfective auxiliary in Flemish dialects obligatorily attracts the participle, perhaps because the perfective feature is strong in the Flemish dialects. The pattern on Map 5 would support such an analysis at first sight, as we see that the order 2–3–1 is also triggered by the presence of a perfective auxiliary and that this phenomenon also occurs in Belgium almost exclusively. However, a comparison of Map 3, 4 and 5 immediately shows that an explanation in terms of parametrisation of obligatory movement and/or strong features is too simple. If all the Belgian dialects that have 1–3–2 on Map 3 have a strong perfective feature on the perfective auxiliary triggering obligatory movement, all these Belgian dialects should have obligatory 2–3–1 on Map 4 and 5. However, Map 4 shows that many dialects in West-Vlaanderen and Vlaams Limburg have 1–2–3 as the only or as an alternative order. On Map 5 the dialects in these provinces only have 1–2–3. For this observation as well, a sociolinguistic analysis is called for.

Finally, the hypothesis that all orders allowed by the grammatical system are in principle allowed in each dialect in the Dutch language area, raises the question as to why the number of attestations of each order differs considerably. In addition to sociolinguistic factors there may be a grammatical factor at play here. The only order that does not involve movement is the 1–2–3 order. If movement is costly, 1–2–3 is the 'cheapest' order and may be expected to be

the unmarked order. Indeed, the overwhelming majority of locations have a 1–2–3 order for the cluster types Mod-Mod-V and Aux-Mod/Asp-V, often as the only option. However, in clusters of the type Mod-Aux-V the orders 1–3–2 and 3–1–2 are more common then 1–2–3. This could mean that sociolinguistic factors can be stronger than economy principles. I leave this issue for further research.

6.   Conclusion

I have argued that language-internal properties determine the variation space in three-verb clusters. In all language varieties considered the 2–1–3 order is systematically excluded, due to the requirement that the original embedding relations be preserved. The order 2–3–1 is systematically excluded in all dialects in hierarchies other than Aux-Mod/Asp-V, due to the lack of Agreement between V1 and V2. The order 3–1–2 is impossible in clusters of the type Aux-Mod/Asp-V, due to the lack of agreement between V1 and V3. Thus, varieties in the Dutch language area do not differ with respect to orders in verb clusters that are clearly ungrammatical.

In the proposed analysis, word order variation in verb clusters derives from optional leftward VP-movement. It was shown that an analysis without movement cannot derive the pervasive 3–1–2 order. If the analysis provided in this paper is correct then it shows that not all variation and optionality can be reduced to morphosyntactic or spell-out properties. Contra current minimalist views, optionality in this paper is considered to be an inherent property of the grammatical system. The system allows a number of syntactically equivalent structures and these orders are, in principle, available in all varieties of Dutch. Speakers and communities pick their choice from these orders.

Finally, it was argued that there are certain patterns in individual and geographic variation about which generative linguistics has nothing to say. That is where sociolinguistics comes in.

Notes

respective audiences, three anonymous reviewers, the editors of this volume and Marcel den Dikken for useful comments and suggestions. Remaining errors are mine.

1. The hierarchically highest verb has index 1, the second highest verb has index 2 and the main verb has index 3.

2. The optionality of PP extraposition (PP occurring in a position following the right-peripheral verb) and scrambling (DP occurring in a position preceding the adverb) are illustrated in (i-a–b) below. Cf. Koster (1974) and Barbiers (1995) for a demonstration of the language internal word order variation that is possible when there is more than one PP.

(i)   a.   Jan heeft <over zichzelf> gesproken <over zichzelf>.
             Jan has   about himself talked   about himself
             "John talked about himself."
     b.   Marie heeft <dat boek> gisteren   <dat boek> gelezen.
             Mary has   that book yesterday that book   read
             "Mary read that book yesterday."

3. I distinguish between judgement tasks and language use because it is an open question as to whether a speaker actually uses all the orders that he claims to occur in his dialect. It is conceivable that a distinction must be made between preferred and possible orders. To be able to answer this question a sufficiently large corpus of spontaneous speech is needed for each dialect such that language use can be compared to the elicited data discussed in this paper.

4. The Syntactic Atlas of the Dutch Dialects (SAND; 2000–2004) is a large-scale project exploring some 120 syntactic variables in 267 dialects in the Dutch language area, which includes The Netherlands and the Dutch speaking part of Belgium. The goal of the project is to publish a two-volume atlas. Volume I (Barbiers et al. 2005) provides maps and descriptions of variation in the syntax of subject pronouns, complementiser agreement, Wh-questions, relative clauses and anaphors. Volume II will involve variation in the syntax of verbs and negation. In addition, a SAND-database is made available on the web containing the spoken, transcribed and partially tagged versions of 267 interviews and the answers to 368 written questionnaires. The database comes with a user-friendly interface and cartographic software to visualise the distribution of syntactic variables. The following institutes participated in the project: Meertens Instituut, the Universities of Amsterdam, Leiden, Gent and Antwerp and the Fryske Akademy. The project was sponsored by VNC (Flemish-Dutch Committee for Dutch Language and Literature) and the Meertens Institute Amsterdam. More information can be found at: http://www.meertens.nl/projecten/sand/sandeng.html. Cf. Barbiers et al. (to appear) for a detailed description of the project.

5. The relevance of the fourth type, 1Aux-2Mod-3V, will be discussed below.

6. The Zuiderzee is located between the provinces of Noord-Holland, Friesland and Overijssel (cf. Map 1 for these provinces). It currently consists of the Flevopolders (southern and eastern part of the former Zuiderzee) and the IJsselmeer (western and northern part of the former Zuiderzee).

7. In a survey of verb cluster orders in German dialects, Schmid and Vogel (2002) claim that 2–1–3 occurs in the Sankt Gallen and Reiderlander Platt dialects under highly marked

conditions involving focus. Given these marked conditions and the fact that these two are the only dialects in a very large sample of German and Dutch dialects that have been reported to have 2–1–3 orders, I assume that they are irrelevant for the present discussion.

8. Note that the number of occurrences of 3V-1Aux-2Asp in previous versions of this paper was claimed to be 8 instead of 3. However, closer examination of the data shows that 3 of these 8 occurrences are not real, they involve false codings. In 2 other locations, 3V-1Aux-2Asp was translated as an absentive with an additional directive particle AWAY. In Standard Dutch, absentive constructions allow a 3V-1Aux-2Asp order as well, as opposed to ordinary verb clusters:

(i)   dat   hij *(uit) wandelen is gegaan
      that he (out)   walking   is gone
      "that he went out for a walk"

In fact, the number of occurrences of 3–1–2 in Aux-Asp-V may be even lower than the remaining 3, since 2 of them occur in the area (in Overijssel) where the absentive construction was found. It may be that these informants have taken the verb cluster as an absentive construction, leaving out the directional particle. This could also be the explanation for the 4 spontaneous occurrences of 3V-1Aux-2Mod in the translation task. In encountering an ungrammatical construction the informants may have changed the construction into an absentive.

9. Despite the large proportion of locations that have 3–1–2 in clusters of the type Mod1-Mod2-V3 (about 25% both in the written and oral interviews) the status of this order in this type is somewhat unclear, in view of the response to the sentence in (i) that was offered during a telephone interview with the same informants that were involved in the oral interviews.

(i)   Ik vind dat iedereen de foto   zien   moet kunnen.
      I   find that everyone the picture see.INF must can.INF
      "I think that everyone should be able to see the picture".

This sentence was reported to occur in only 4 out of 245 locations (1.6%). A possible explanation is that telephone interviews yield less reliable results than written or oral interviews. Another possibility is that there is an interfering linguistic factor. Certain main verbs, at least in Standard Dutch, can (marginally) leave the cluster by focus movement, whereas others cannot, for reasons that are unclear. Thus, whereas (ii) is impossible, focalisation on the basis of the test sentence from the oral interview is possible (iii). The status of the test sentence from the written questionnaire is somewhere in between (iv):

(ii) *Ik vind dat iedereen de foto   zien   wel moet kunnen.
       I   find that everyone the picture see.INF affir must can.INF.

(iii) ?Ik vind dat iedereen zwemmen wel   moet kunnen.
       I   find that everyone swim.INF affir must can.INF

(iv) ??Ik weet dat Jan hard werken   wel   moet kunnen
        I   know that John hard work.INF affir must can.INF

Although the explanation for the low incidence of the 3–1–2 order in (ii) as compared to the sentences from the written and oral interviews requires further research, it seems fair to conclude from the marginality of (iv) that the high incidence of 3–1–2 orders in the results of the written questionnaire cannot be completely be attributed to the possibility of focus movement. Put differently, I conclude that the order 3–1–2 is a possible order in clusters of this type. See also Den Besten and Broekhuis (1989) and Broekhuis (1997) for discussion of this issue.

10. Note that the symbol '+' on the map is a combination of a vertical and a horizontal line. These locations have both 3–1–2 and 1–3–2 orders.

11. A combination of a vertical and a horizontal line yields '+'.

12. In some of these dialects, in particular in Friesland, the equivalent of *gaan* 'go' requires a verbal complement introduced by *te* 'to'. As far as I can see, this does not have consequences for the word orders in the cluster here.

13. In this paper I abstract away from the position of non-verbal arguments and adjuncts.

14. For the purpose of this paper I define c-command as follows (cf. Kayne 1994 for the full definition):

X c-commands Y if all nodes dominating X dominate Y. Asymmetric c-command holds between X and Y if X c-commands Y but Y does not c-command X. For example, in (4a) V1 c-commands V2 because VP1 dominates V1 and VP1 also dominates V2. However, V2 does not c-command V1, since V2 is dominated by VP1 and VP2, and VP2 does not dominate V1. Consequently, V1 asymmetrically c-commands V2. According to Kayne (1994), if V1 asymmetrically c-commands V2, V1 linearly precedes V2. If we apply these definitions to the structures in (4) we see that the different ordering of the nodes does not have consequences for the resulting linear ordering. In all cases V1 asymmetrically c-commands V2 and V2 asymmetrically c-commands V3, yielding the linear order V1-V2-V3.

15. The original position of a moved constituent is indicated by a strike-through.

16. Cf. Broekhuis (1997) and Haegeman (1998), among many others, for leftward VP-movement accounts of word order variation in verb clusters. In their analyses, VP moves into the Spec of a functional projection. Broekhuis (1997) claims that 3–1–2 is impossible for clusters of the type Modal-Modal-Verb and rules out the derivation in (6e).

17. Koopman and Szabolcsi (2000) propose an antisymmetric analysis of verb cluster orders that is crucially different from the one proposed here. Adopting the minimalist program, they assume that movement is always obligatory. If nothing more were said, the only order that could be derived is 3–2–1. To derive the other orders while keeping movement obligatory, they introduce the mechanism of remnant roll up. The basic idea is that it is possible to move a constituent XP out of a constituent YP and then move YP containing the trace of XP across the landing site of XP (cf. (i-a–c). This type of derivation is independently motivated for VP-topicalisation (i-d, e); cf. Den Besten & Webelhuth 1990). If objects are generated within the VP and if the well-known generalization that heads cannot be topicalised in Dutch is correct, (i-d) must be analysed as in (i-e), i.e. first the object moves out of VP, then the VP, containing the copy of the moved object, is topicalised.

(i)   a. $[_{VP1}$ V1 $[_{VP2}$ V2 $[_{VP3}$ V3 $]]]$ Move VP3 across VP2

      b. $[_{VP1}$ V1 $[_{XP}$ $[_{VP3}$ V3$]$ $[_{VP2}$ V2 $[_{VP3}$ ~~V3~~ $]]]$ Move remnant VP2 across V3 and V1

      c. $[_{ZP}$ $[_{VP2}$ V2 $[_{VP3}$ ~~V3~~ $]]$ $[_{VP1}$ V1 $[_{XP}$ $[_{VP3}$ V3$]$ $[_{VP2}$ ~~V2~~ $[_{VP3}$ ~~V3~~ $]]]]]$

      d. $[_{VP}$ Gelezen] heeft Jan   het boek niet $[_{VP}$ ~~gelezen~~]

           Read   has John the book not   read

      e. $[_{VP}$ [~~het boek~~] gelezen] heeft Jan   het boek niet $[_{VP}$ [~~het boek~~] gelezen]

           the book read   has John the book not   the   book read

Although analyses of this type look maximally restrictive in that they only allow overt left-ward XP-movement, they derive all six logically possible word orders in three-verb clusters (cf. Barbiers 2003). In particular, as (ii-a–c) show, it is possible to derive the 2–1–3 or-der which is systematically absent in the Dutch and German dialects. In Koopman and Szabolcsi (2000) such orders are ruled out by complexity filters, but since these filters are parametrised, it is completely accidental that there do not seem to be languages that allow the 2–1–3 order. The optional leftward VP-movement analysis derives the ungrammaticality of this order automatically as a fundamental property of the system.

18. Cf. Zwart (1995) for an alternative analysis. Zwart assumes that movement of infinitivals involves head movement whereas movement of participles involves movement of maximal projections. The fact that particles can not be stranded when an infinitival moves, whereas they can be stranded when a finite verb moves (cf. example (7)) casts doubt on such an analysis. It is also not clear why infinitivals and participles would differ from each other with respect to the X-bar level that is attracted. The present proposal is based on the null hypothesis that as far as syntax is concerned infinitives and participles only differ in their morphosyntactic feature specification.

19. The exact label of the feature is immaterial. Any feature that distinguishes between modal/aspectual verbs and main verbs will do.

20. This situation should be distinguished from the one in which VP2 is pied-piped as a result of agreement between V1 and VP3 in SpecVP2 (Option 2b). This will always give rise to the order 3–2–1. When VP3 does not move to SpecVP2, agreement between V1 and VP3 can trigger movement of VP3 to SpecVP1, but in such cases VP3 cannot pied-pipe VP2. We seem to be dealing with a general contrast between specifiers and complements here. For example, whereas a Wh-element in SpecDP can, and in some cases must pied-pipe the full DP that contains it (i)–(ii), a Wh-element as a complement of N cannot (iii):

(i) $[_{DP}$ Wiens boek] heb je   bestudeerd?

      whose book   have you studied

(ii) $^{\star}[_{DP}$ Een boek van wie] heb je   bestudeerd?

      a   book of who have you studied

(iii) Van wie heb je   een boek bestudeerd?

      of who have you a   book studied

21. EPP stands for Extended Projection Principle. The Projection Principle (Chomsky 1981) states that all arguments of a predicate must be projected in syntactic structure. However, this principle does not capture the fact that every sentence must have a subject, since the subject need not be an argument of a predicate, as in the case of expletive subjects. The Ex-

tended Projection Principle (EPP) states that every sentence must have a subject. Chomsky (2001) assumes that this requirement derives from the presence of an EPP-feature and generalises this to all cases of movement, assuming that there is always an abstract EPP-feature triggering movement.

22. In addition to Dutch, Frisian has the status of an official language in the Netherlands, as opposed to all other varieties of Dutch in the Netherlands. As a result, Frisian is taught in schools, there are grammars and dictionaries of Frisian and there is a certain normative pressure.

References

Baker, Mark (1996). *The Polysynthesis Parameter*. Oxford Studies in Comparative Syntax. New York/Oxford: Oxford University Press.

Barbiers, Sjef (1995). The Syntax of Interpretation. Ph.D. dissertation, Leiden University (= *HIL Dissertation* 14).

Barbiers, Sjef (2003). "Book Review of Hilda Koopman & Anna Szabolcsi (2000) *Verbal Complexes*". *The Journal of Comparative Germanic Linguistics, 6*(1), 53–79.

Barbiers, Sjef (2005). "Variation in the Morphosyntax of ONE". *Journal of Comparative Germanic Linguistics.*

Barbiers, Sjef, Hans Bennis, Gunther De Vogelaer, Magda Devos, & Margreet van der Ham (2005). *Syntactic Atlas of the Dutch Dialects* (*SAND), Volume 1*. Amsterdam: Amsterdam University Press.

Barbiers, Sjef, Leonie Cornips, & Jan Pieter Kunst (to appear). "The Syntactic Atlas of the Dutch Dialects (SAND). A Corpus of Elicited Speech as an On-line Dynamic Atlas". In Joan C. Beal, Karen P. Corrigan, & Hermann Moisl (Eds.), *Using Unconventional Digital Corpora: Synchronic Corpora* 1. London: Palgrave Macmillan.

Besten, Hans den & Gert Webelhuth (1990). "Stranding". In Gunther Grewendorf & Wolfgang Sternefeld (Eds.), *Scrambling and Barriers* (pp. 77–92). Amsterdam: John Benjamins.

Besten, Hans den & Hans Broekhuis (1989). "Woordvolgorde in de Werkwoordelijke Eindreeks". *GLOT, 12*, 79–137.

Bloemhoff, Hein (1979). "Heranalyse van een Stellingwerver Oppervlaktestructuur". *Us Wurk, 28*, 31–38.

Broekhuis, Hans (1997). "Nogmaals Verb Projection Raising". *TABU, 27*(1), 1–27.

Chomsky, Noam (1981). *Lectures on Government and Binding*. (= *Studies in Generative Grammar, 9.*) Dordrecht: Foris Publications.

Chomsky, Noam (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam (2001). "Derivation by Phase". In M. Kenstowicz (Ed.), *Ken Hale. A Life in Language* (pp. 1–52). Cambridge, MA: MIT Press.

Cornips, Leonie & Willy Jongenburger (2001). "Elicitation Techniques in a Dutch Syntactic Atlas Project". In H. Broekhuis & T. van der Wouden (Eds.), *Linguistics in the Netherlands* 2001 (pp. 57–69), 18. Amsterdam/Philadelphia: John Benjamins.

Cornips, Leonie & Cecilia Poletto (2005). "On Standardising Syntactic Elicitation Techniques". *Lingua, 115*(7), 939–957.

Den Dikken, Marcel (1994). "Minimalist Verb Projection Raising". In C.-J. W. Zwart (Ed.), *Minimalism and Kayne's Antisymmetry Hypothesis*. (= *Groninger Arbeiten zur Germanistischen Linguistik, 37*, 71–88.)

Den Dikken, Marcel & Eric Hoekstra (1997). "Parasitic Participles". *Linguistics, 35*, 1057–1089.

Evers, Arnold (1975). The Transformational Cycle in Dutch and German. Ph.D. dissertation, Utrecht University.

Haegeman, Liliane (1998). "Verb Movement in Embedded Clauses in West Flemish". *Linguistic Inquiry, 29*(4), 631–656.

Henry, Alison (1995). *Belfast English and Standard English: Dialectal Variation and Parameter Setting*. New York: Oxford University Press.

Henry, Alison (2002). "Variation and Syntactic Theory". In J. K. Chambers et al. (Eds.), *The Handbook of Language Variation and Change* (pp. 267–282). Malden/Oxford: Blackwell.

Hinterhölzl, Roland (1998). Restructuring as XP-Movement. Ph.D. dissertation, University of Southern California.

IJbema, Aniek (2002). Grammaticalization and Infinitival Complements in Dutch. Ph.D. Dissertation, Leiden University. (= *LOT dissertation 54.*)

Kayne, Richard (1994). *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

Koopman, Hilda & Anna Szabolcsi (2000). *Verbal Complexes.* Cambridge, MA: MIT Press.

Koster, Jan (1974). "Het Werkwoord als Spiegelcentrum". *Spektator, 3*, 601–618.

Kroch, Anthony (1994). "Morphosyntactic Variation". In K. Beals (Ed.), *Papers from the 30th Regional Meeting of the Chicago Linguistics Society*. Chicago: CLS.

Pesetsky, David & Esther Torrego (2001). "T-to-C Movement: Causes and Consequences". In M. Kenstowicz (Ed.), *Ken Hale. A Life in Language* (pp. 355–426). Cambridge, MA: MIT Press.

Preston, Dennis (2004). "Language with an Attitude." In J. K. Chambers et al. (Eds.), *The Handbook of Language Variation and Change* (pp. 40–66). Oxford: Blackwell.

Schmid, Tanja & Ralf Vogel (2002). "Dialectal Variation in German 3-Verb Clusters". Paper presented at *GLOW*, Amsterdam, 2002.

Stroop, Jan (1970). "Systeem in Gesproken Werkwoordsgroepen". In J. Stroop (Ed.), *Nederlands dialectonderzoek*. Amsterdam: Huis aan de drie grachten.

Zuckerman, Shalom (2001). The Acquisition of "Optional" Movement. Ph.D. dissertation, University of Groningen. (= *Groningen Dissertations in Linguistics* 34.)

Zwart, C. Jan-Wouter (1995). "A Note on Verb Clusters in the Stellingwerf Dialect". In M. den Dikken & K. Hengeveld (Eds.), *Linguistics in the Netherlands 1995*. Amsterdam/Philadelphia: John Benjamins.

The third dimension of person features*

Paola Benincà and Cecilia Poletto

Università di Padova / Istituto di Scienze e Tecnologie
della Cognizione CNR

1. Introduction

In recent literature much work has been devoted to understanding how agreement and person systems work (see among others Bianchi 2004; Sigurdsson 2004). The present article intends to re-visit this problem exploiting the mass of data accumulated by descriptive and traditional etymological research on inflectional endings and on analogy inside the verbal and pronominal paradigms in Romance. Our aim is thus to provide a model of how person features are organized in the grammar by uncovering the paths through which analogy can extend pronominal forms across paradigms.

Analogy in its traditional meaning is best conceived of as a cover term for a number of different phenomena. Here, we intend to examine a proper subset of these, namely, those cases of extensions of a form that we believe to be motivated by semantic similarities. This will enable us to provide both a sketch of how some analogical processes develop and to construct a first representation of the category "person", which, as already proposed in much recent work, is a complex object both in the morphological and in the syntactic component. We will use mainly data from the pronominal domain, and refer to the verbal paradigm only occasionally. There is a substantive reason for this procedure: while the pronominal paradigms in several Romance languages display and dialects a number of extensions, which can be studied and ordered into an implicational scale, the verbal paradigms do not provide us with the same amount of empirical evidence, because personal endings are in general very well preserved with respect to their Latin original forms. Only the present subjunctive shows an extension in the singular persons, which means mainly a loss

of personal distinctions. Analogical extensions go – as a general pattern – from one conjugation to another, from one tense to another, and in general not from one person to another. Nevertheless, some particular cases will be mentioned occasionally.

On a more general basis, we will show how detailed data drawn from very closely related Romance dialects can provide a refined instrument of inquiry, which goes at least as deep as typological comparison among a wide range of non genetically-related languages. Although many features considered in the typological and generative literature on person deal with categories like dual, trial etc., which do not feature prominently in our inquiry, the main objective of this chapter is to motivate an analysis that explains, in a rigorous manner, the interesting connections between different persons by investigating cases where the same form is used for different persons in Romance. More specifically, the description and analysis of extension patterns of single forms highlights the crucial status of fourth and fifth person, which very often constitute the "bridge" for moving from the system of non-deictic (third persons) to the system of +deictic (first and second person) or vice-versa.[1] Moreover, a structural representation in terms of a tree which shares the same properties as syntactic trees will be shown to be untenable. We propose, instead, a three-dimensional structure where nodes can be activated or not, but no node can be marked with a negative value. This system, we believe, accounts for a number of extension patterns, first of all for the reflexive clitic form *s+V*, and for vocalic subject clitics, but also for many other cases.

Although, our proposal will no doubt have to be fine-tuned when applied to other language groups, because it only considers a single language group, and surely categories like dual, trial etc. will also have to be added to the representation we provide here. It constitutes a starting-point for bringing the traditional concept of analogy under the auspices of a more principled framework.

The chapter is organized as follows: in Section 2 we show some analogical extensions from one person to another with regard to subject and reflexive clitics, which seem to follow the same path, though in opposite directions. Section 3 offers a sketch of the manner in which person features in Romance are structured, showing that a third dimension is needed to capture the analogical extension patterns described in Section 2. Section 4 provides further empirical evidence in favour of the three-dimensional structure we propose for person features and focuses on the verbal paradigm, object clitics and pronouns.

2.   The general perspective

The set of data we investigate here comes from the comparison of the very closely related systems of Northern Italian dialects are. What we present as extension patterns of a single form that for instance can be used in dialect A only for first person and in dialect B for first and fourth person is not a model of coexisting systems in the mind of a single speaker, in the sense that no speaker has both dialect A and dialect B in his mind (and if so, this is only by chance). The type of variation that is considered here is only geographical and represents a range of minimally differing systems in a linguistically unitary area; it is not to be intended in the sense of variation inside the same speaker, where the same speaker masters more than one grammar (or has two alternative rules for the same phenomenon). Hence, this work does not investigate the problem of how different grammars relate inside the competence of a single speaker. It seeks for an explanation of why dialectal variation is as it is, in other words, why there are implicational scales and dialectal variation (DV) is not random. Finding an algorithm behind DV has far reaching consequences, as it shows on the one hand that dialectal variation is fundamentally different from typological variation and, on the other, it leads us to postulate that dialects of a coherent geographical space have the same underlying system (or diasystem) and that DV is the reflex of shallow differences.

DV might instead reflect the single steps of diachronic change, where a system evolves into another through a series of logically ordered procedures. Although, due to lack of relevant and sufficient text sources of Old NIDs, it is impossible to prove for the case in question that the diachronic path has been the same in all dialects; nevertheless DV might be enlightening for constructing hypotheses on language change as well. If this view turns out to be correct, language change should not be seen as a "catastrophic" and abrupt mutation, but as a dialect evolving into another, even if written records only show a change when it has become important and steady.

Leaving the speculations on the relation between DV and diachronic change aside, in some sense we could even say that this work is not concerned with variation at all but only with the comparison of different systems. Nevertheless, we believe that this type of comparison is illuminating because, coming from minimally different and genetically-related languages, it provides the range of the possible values that a pronominal form can assume, and the ordering in which it extends its value from one person to another can give us a detailed picture of which persons are more closely related. By analysing the ordering of extensions of a single form across different dialects we can

hypothesize which are the basic components (from now on, features) of the persons of the Romance paradigm. This is the basic ultimate goal of this work, which we set into a formal proposal, namely the one put forth in Section 4 as a three-dimensional structure. Considering geographical DV we hope to create a connection between the modern and the Latin system, or at least to reconstruct some aspects of it, given that simplification might have rendered unrecoverable parts of the initial system.

One can conceive this work as similar to the work of a phonologist trying to identify what the distinctive phonological features in a given language are. What we are trying to find out here are the basic distinctive semantic features, i.e. the minimal units which associate to form the six persons of the Romance paradigm. This can only be seen as a first step towards the far more complex task of hypothesizing the universal inventory of semantic features that can make up a person; this problem has already been addressed by Harley and Ritter (2002) on the basis of various languages belonging to different families; their hypothesis is not yet directly comparable with ours, as the data we are dealing with are too detailed with respect to those that Harley and Ritter arrived at on the basis of the phenomena they considered; this is a sign that more empirical work is needed before the dialectological and the typological perspective can meet. The type of data that concerns extension pattern among closely related languages cannot be inserted into a typological perspective, but could obviously be replicated for other language groups once a sufficient number of dialects are investigated.[2, 3]

A note on how the data lying at the basis of this research have been gathered is in order. Our primary source is the ASIS data base (http/:asis-cnr.unipd.it), to which we refer for the inquiry protocol and the methodology (see also Cornips & Poletto 2005). We also used the AIS Dialect Atlas of Italy and a number of secondary sources (descriptive grammars and texts) as a control sample to test our hypothesis on the extension patterns. The reason why we think that the latter are also valid sources is that, in our experience, morphology appears to be a very stable component of language even in situations of bi- or multilingualism (this fact is well known among dialectologists and historical and general linguist: see among many others Weinreich 1981); the inspection of morphological features is in fact what is often used as a device for distinguishing between one language and another in situations of language contact. Granted this, it is clear why it does not represent a problem during elicitation tests, so it does not require any special techniques to be obtained. Therefore, we believe our choice of using secondary sources as a control sample for our hypotheses is justified.

2.1 Setting the problem: *Se* versus *le*

In general, the set of reflexive clitic pronouns in the Romance languages distinguishes between the third person pronoun *si/se/sa* (hereafter *s*+V (or V+*s*)), which is used for the masculine and feminine third person singular and plural as well as first and second person pronouns.[4, 5] We will start by examining some aspects of Kayne's (2002) hypothesis on reflexive and non-reflexive clitic forms that are relevant for the present work. Kayne (2002) points out that no Romance language seems to distinguish singular and plural in the reflexive form of third person. The following paradigm of the reflexive clitic forms in the most well-known Romance languages illustrates this point (Table 1).

Kayne (2002) shows that the vowel is epenthetic in the reflexive clitic form *s*+V, because it always corresponds to the unmarked vowel used in epenthesis in each language. Moreover, he suggests that the third person *s*+V form belongs to a paradigm containing the *m* and *t*, *n* and *v* forms, which are also used for personal pronouns as well as reflexives, and also have an epenthetic vowel. On the contrary, the third person object clitic *l*+V does not feature in the paradigm that contains the first, second, fourth and fifth person forms *m*, *t*, *n* and *v*.[6] Kayne hypothesizes that this is so because the non reflexive third person clitic *l*+V is bymorphemic: in the form *l*+V, the vowel is an agreement marker displaying gender and number features, while this is not the case for the *s*+V and *m,t,n,v*+V forms.

Before illustrating the extension pattern of the *s*+V form, it is worth noting that Kayne's hypothesis concerning the fact that *s*+V is in a single paradigm with the *m,t,n,v,*+V forms, while *l*+agreement is excluded from it, is confirmed by the fact that the extensions found in Romance languages involve the *s*+V form frequently substituting for the *m,t,n,v,*+V forms, but never for the *l*+V forms.

Although the pattern in Table 1 looks rather homogeneous, it does not take into consideration, in fact, a number of non standard varieties of Romance

Table 1.

	Italian	French	Spanish	Portuguese	Rumanian	Catalan
1. pers.	mi	me	me	me	mă	em
2. pers.	ti	te	te	te	te	et
3. pers.	si	se	se	se	se	es
4. pers.	ci	nous	nos	nos	ne	ens
5. pers.	vi	vous	os	vos	vă	us
6. pers.	si	se	se	se	se	es

that extend the form *s+V* to other persons; the extension seems to follow an implicational scale.

As a first step towards understanding the extension pattern associated with reflexive clitics and the implicational scale it gives rise to, consider the data in (2) below, which are taken from Venetian. The pattern of particular interest in this case is as follows: if a language uses the *s+V* form for another person of the paradigm, this is the first person plural or, as we prefer to call it, the fourth person. This is a strategy that is regularly employed in several varieties of Veneto; the data used here come from the dialect spoken in Venice:[7]

(2) a. 1. pers. Me   go   meso i   calseti
               myself have put   the socks
               "I have put my socks on"

     b. 2. pers. Ti te   ga   meso i   calseti
               you yourself have put   the socks
               "You have put your socks on"

     c. 3. pers. El **se**   ga   meso i   calzeti
               he himself has   put   the socks
               "He has put his socks on"

     d. 4. pers. **Se**   gavemo meso i   calseti
               ourselves have   put   the socks
               "We have put our socks on"

     e. 5. pers. Ve   gavè meso i   calseti
               yourself have put   the socks
               "You have put your socks on"

     f. 6. pers. I   **se**   ga   meso i   calseti
               they themselves have put   the socks
               "They have put their socks on"

Indeed, this extension is widespread not only in the Northern Italian domain, but also in Southern Italy, where cases similar to (2) are also attested (albeit in a somewhat scattered fashion) in dialects spoken in the regions of Lazio, Umbria, Northern Abruzzo and Southern Basilicata (see AIS charts IV 660 *ci laviamo* 'we wash ourselves' and VIII *spicciatevi* 'hurry up') while the form *n+V* is widespread throughout Sicily, Northern Basilicata, Puglia, and Southern Campania.[8]

The second extension pattern relevant to this discussion involves the second person plural (fifth person), and is illustrated by the data in (3) below from the variety of Rodoretto di Prali (in Western Piedmont, a Franco-Provençal dialect):

(3) a. 1. pers. Me   seou sta
              myself am   sat
              "I sat down"

     b. 2. pers. Tu t   se sta
              you yourself are sat
              "You sat down"

     c. 3. pers. A **s'**   è sta
              he himself is sat
              "He sat down"

     d. 4. pers. Nou **s'**   soun (e)sta
              we   ourselves are   sat
              "We sat down"

     e. 5. pers. Ou **s'**   se sta
              you yourselves are sat
              "You sat down"

     f. 6. pers. I   **s'**   soun (e)sta
              they themselves are   sat
              "They sat down"

Given that the extension runs along an implicational scale, the extension step to the fifth person also implies that the *s+V* form has been extended to fourth person as well.

In some other dialects the two forms with *v+V* and *s+V* alternate depending on the syntactic context: the *s* form being preferred with imperatives and/or in main interrogative clauses while the *v* form is used in all other contexts:

(4) a. Z   maria pa?   *Maddalena* (*Piedmont*)
         REFL marry not?
         "Aren't they going to marry?"

     b. Vü vi   sumè
         you REFL stand up
         "You stand up"

(5) a. spostesse!   *Sacile* (*Friulian*)
         move-yourselves!
         "Move!?"

     b. ve   spostè
         yourselves move
         "You move"

Another interesting case is provided by those dialects (described by Benincà & Vanelli 1982) in which the *s* form alternates with a null form.[9]

(6) a. Si   seso   vistús?   *Moimacco* (*Friulian*)
         yourself are-you dressed?
         "Did you put your clothes on?"
     b. e   sus sintàs
         you are seated
         "You sat down"

Examples (4) and (5) above show that these extension patterns are not only sensitive to morphology (occurring as they do only within what is analysed by the speakers as a paradigm) but that they are also dependent on syntactic context.[10]

   Other dialect systems within the Romance group show the reflexive form *s+V* for third, fifth and, interestingly, also for second person:[11]

(7) a. 3. pers. El s'   è metù i   calzet bianch
                   he himself is put the white socks
                   "He has put on white socks"   *Montagnola* (*Lombard*)
     b. 6. pers. I s'   en metù i   calzet bianch
                   they themselves are put   the white socks
                   "They have put on white socks"
     c. 5. pers. S   i metù i   calzet bianch
                   yourselves is put   the white socks
                   "You have put on white socks"
     d. 2. pers. Te s'   e metuda i   calzet bianch
                   you yourself is put   the white socks
                   "You have put on white socks"

(8) a. 1. pers. ma   sum setàa giò   *Lugano* (*Lombard*)
                   myself are   sit   down
                   "I sat down"
     b. 2. pers. ta   sa   set setàa giò
                   you yourself are sit   down
                   "You sat down"
     c. 3. pers. al s'   è setàa giò
                   he himself is sit   down
                   "He sat down"
     d. 4. pers. (o) sa   sem setàa giò
                   (we) ourselves are sit   down
                   "We sat down"
     e. 5. pers. (va) sa   sii setàa giò
                   yourself+yourself are sit   down

    f.   6. pers. i   s'   è   setàa giò
                    they themselves are sit   down
                    "They sat down"

In yet others (such as Lugano in (9) below), the fifth person, the second and the first display split forms containing two clitics:

  (9) a.   1. pers. ma   sa   *Lugano* (*Lombard*)
                 myself
       b.   2. pers. (ta) sa
                 yourself

       c.   5. pers. (va) sa
                 yourselves

While second and fifth person optionally alternate with non-doubling forms in which only *s+V* occur, the first person does not have an *s+V* form in any of the dialects present in our data base, i.e. their occurrence is entirely restricted to the doubled form, as in example (9a). These are probably best analysed as a splitting of the features, such that the *m/t/v* form represents the person feature while the *s* form represents the reflexive feature. If this lack of a non-split form for the first person is really confirmed by further investigation of additional dialectal materials, i.e. it is not simply a consequence of our particular data base, which might be skewed in this respect; this may well indicate that the *s+V* form, although compatible with the first person because of splitting, cannot express the features of a first person.

    On the basis of the data illustrated in (2) through (9), we can conclude that the extension pattern observed for reflexive *s+V* forms is the one illustrated in Figure 1 (the first person singular being bracketed in the schema because only doubled forms are actually attested).[12]

    third person → fourth person → fifth person → second person → (first person)

Figure 1.

2.2 Further refining the problem: Vocalic clitics

In this section, we will restrict our discussion to cases of so-called vocalic clitics.[13]

    The pattern of extension in this case is as follows: etymologically, the vocalic forms seem to derive from the first person singular pronoun 'EGO'.

Although this view is not shared by all scholars, there is at least one empirical argument in favour of this idea and it concerns evidence from French dialects in which the unambiguous clitic form *je* 'I' extends to fourth person. In Gruyère, for instance, the form corresponding to standard French 'je' and 'nous' (we) are both 'j' (cf. Shlonsky & De Crousaz 2002; Kayne 2002). A similar pattern also occurs in some NIDs, such as the Ligurian dialect of Oneglia (but the pattern can also be found in a scattered fashion throughout the geographical domain under investigation), in which only first and fourth persons display a vocalic clitic:

(10)   a.   **A**   mangiu   *Oneglia* (*Ligurian*)
            I   eat
       b.   **A**   mangiammu
            we eat

The pattern then extends to the fifth person as in (11) and (12) below; in other dialects, such as Loreo in (13), a further extension to the second person can also be identified:[14, 15]

(11)   a.   **E**   no   podeva tior   (*Calmo 66*) *Venetian of the XVI century*
            I   not could   take
            "I could not take"
       b.   **E**   no   se   inganemo   (*Calmo 66*)
            we not ourselves mistake
            "We are not wrong"
       c.   **E**   no   podé   (*Calmo 66*)
            You not can
            "You cannot"

(12)   a.   **A**   magn   *Bologna*
            I   eat
       b.   **A**   magnén
            we eat
       c.   **A**   magnè
            You eat

(13)   a.   1. pers. **A**   magno   *Loreo* (*Veneto*)
                      SCL eat
                      "I eat"
       b.   2. pers. **A**   te   magni
                      SCL SCL eat
                      "You eat"

   c. 3. pers. El/la magna
                 SCL   eat
                 "He/she eats"

   d. 4. pers. **A**   magnemo
                 SCL eat
                 "We eat"

   e. 5. pers. **A**   magnè
                 SCL eat
                 "You eat"

   f. 6. pers. I/le magna
                 SCL eat
                 "They eat"

Indeed, there are yet more non-standard varieties that have a vocalic clitic for all persons; in these dialects the clitic presents two forms, one being used for first, second, fourth and fifth person and another which marks third persons. We consider these cases as being similar to the ones in (13) immediately above and (14) below since the distinction demarcates third persons from all of the others.

(14) a. I   mangi   *S. Michele al T.* (*Friulian*)
          SCL eat
          "I eat"

      b. I   ti   mangis
          SCL SCL eat
          "You eat"

      c. A   l   mangia
          SCL SCL eat
          "He eats"

      d. I   mangin
          SCL eat
          "We eat"

      e. I   mangè
          SCL eat
          "You eat"

      f. A   mangin
          SCL eat
          "They eat"

This last extension pattern includes third person as well; the Lombard dialects, in particular, (as (15) demonstrates) are unique in having a discrete form that acts as a marker for all persons. It is noteworthy that there is no Italian dialect

where the vocalic form extends to third singular but not to third plural or vice versa. More generally, in this kind of clitics number does not seem to play any role; this lack of sensitivity to number replicates the data that we outlined above for reflexive clitics.

(15) a. 1. pers. A   vegni mi   *Lugano* (*Lombard*) Vassere (1993)
                    SCL come I
                    "I come"

      b. 2. pers. A   ta   vegnat ti
                    SCL SCL come   you
                    "You come"

      c. 3. pers. A   vegn luu
                    SCL come he
                    "He comes"

      d. 4. pers. A   vegnum
                    SCL come
                    "We come"

      e. 5. pers. A   vegnuf
                    SCL come
                    "You come"

      f. 6. pers. A   vegn lur
                    SCL come they
                    "They come"

2.3 Summary of extension patterns

Taking all the evidence presented in §§2.1 and 2.2 into account, the extension pattern of vocalic clitics can be summarized as in Figure 2.

Comparing this schema with that of Figure 1 (replicated above as Figure 3), the two extension patterns of the *s+V* clitic reflexive forms and of vocalic subject clitics overlap in the central domain of the paths, though they are different at the extremes.

first person → **fourth person** → **fifth person** → **second person** → third person

Figure 2.

third person → **fourth person** → **fifth person** → **second person** → (firstperson)

Figure 3.

In order to explain these rather intriguing facts, we will start from the assumption, in the spirit of Poletto (2000) as well as Harley and Ritter (2002), that person consists of more basic components which it shares with other persons.

3.   Restrictions on extension

We can formulate the hypothesis that a given form (either pronominal or verbal) can extend from one person to another when the two persons have at least one component in common. The following descriptive generalization, therefore, captures the schemes predicted in Figures 1 and 2, viz.: 'Extension is possible if and only if the two persons involved share at least one value for one feature in their composition.' This statement accounts for the fact that extension patterns are not always identical, because they can move from one person to all the persons that have at least one basic component in common. Given that the feature composition of each person is complex, we expect to find more than one 'extension path', depending on the feature taken into account by the extension paradigm. As for now, we have found only the extension paradigms illustrated in Figures 2 and 3. As noted, they are identical in their central part, though not at the extremes. The theoretical proposal we put forth in Section 4 represents a formal codification of these two paradigms; if, after more empirical work, it should turn out that there are more, our analysis would obviously have to be modified.

Going back to the two extension patterns illustrated in Figures 2 and 3 above, we can, therefore, infer that fourth person has something in common with both first and third persons, and that second person also has a feature in common with first and third. On the other hand, as we will demonstrate below, when two persons have more features in common, the extension seems to happen more frequently, thus creating privileged paths of extensions, like the one we see in the central domain of the patterns illustrated in the schema above. For instance, fourth and fifth person are singled out as a class by a number of verbal and pronominal extensions which suggest that their feature composition must look very similar.[16] Although a principle like the generalization we proposed above seems to adequately account for the facts, a more principled account of the extension pattern in question will be provided in the next section. In addition, we will turn our attention to formalizing a proposal that more adequately explains the feature composition of the persons in Romance dialects such as those we have already introduced.

Before moving on to this discussion, it is also worth pointing out that there are a number of additional restrictions active on extension patterns. As we have already mentioned, even if on the basis of few cases, a given syntactic context can favour extensions. As we saw with reflexives, the syntactic position of the verb is a crucial factor determining the form of the pronoun used. Thus, the reflexive *s+V* form can surface only when the verb raises to certain positions in the clause. The morphological composition of the forms to which extension applies is an additional factor which can influence the process. This is particularly clear in the case of reflexives examined above. Kayne's (2002) hypothesis that the *s+V* form belongs to the same paradigm as the *m/t/n/v +V* forms while the third person object clitics, on account of their having an agreement morpheme, does not, is confirmed by our data, provided the following natural assumption is made: extension patterns are possible inside the same paradigm, which is established by speakers on the basis of their precise morphological analysis of the form in question.

4.   Person features in three-dimensional space

Our proposal for the feature composition of persons rests on an observation which goes back at least to Benveniste (1966), who noted that a distinction has to be drawn between the deictic persons and the non-deictic persons, namely third persons versus first, second, fourth and fifth person. It is this same distinction that we referred to earlier in the case of Friulian vocalic clitics (see example (14)), which have a form distinguishing third persons from first, second fourth and fifth, namely non-deictic from deictic persons.

We think that the +/− deictic distinction has at its base the same feature that the system of demonstrative pronouns encodes, namely, a contrast between what is present in a conventionally defined space in the domain of the discourse and what is absent from that space. An empirical argument in favour of this view is provided by the etymology of third person pronouns, which are generally derived from the 'non-proximate' demonstrative pronoun, indicating an object located far from the speaker, hence absent from the physical space that includes the speaker. This is true across the entire Romance language family, in which the clitic forms for third person pronouns all originally derive from the Latin pronoun *ille* 'that', this is true for French *le*, Italian *lo*, Spanish *lo*, etc.[17] We propose that the first distinction between third persons and the others is something akin to the semantic notion [here]. A cautionary note is, however, in order in this case, as the definition of what is present or not does

Person

+here +there

+plural

+speaker +hearer
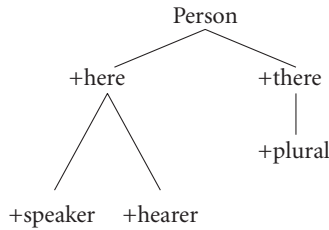
Figure 4.

not always overlap with the actual occurrence of a person in some physical space. For instance, it is possible to refer to a person who is in the same room as the speaker as *he/she*. In other words, under such circumstances they are absent from the fictive space defined in the universe of discourse. Similarly, it is possible for a speaker to employ the pronoun *you* to address another person present in the discourse, although he/she is not, necessarily, in attendance in the same physical space (as would be the case, if the interlocutors were talking to one another by telephone). Hence, what is defined as deictic or not does not always coincide with the real physical position of speakers in discourse. Rather, it should be conceived of as an abstract notion of common space which belongs to the principles setting the universe of discourse.

In addition, among first and second person we also need to draw a distinction in terms of [speaker] and [hearer], with first person being [+speaker] and second person being [+hearer].

From these preliminary distinctions, therefore, a tree like that given in Figure 4 can be conjectured.

The topmost node, i.e. 'person' gathers the feature specifications that are activated for each person. They are [+here] for the deictic persons, and [+there] for the non-deictic ones. Among the latter group, it is also possible to activate the [plural] node, thus deriving a third person plural. Note that if we hypothesize that the Person node is complex, i.e. it inherits only the specifications that are activated inside its tree, there can be no negative specifications, so second person cannot simply be [–speaker], but has to be [+hearer]. In other words, given that only the activation of a node is inherited by the dominating node, there can be no '–' in the system.

Furthermore, there is also an empirical argument for not choosing the possibility of negative settings of a feature, where, for instance, the hearer is derived through a negative setting of the [+/–speaker] feature. If we assume that fourth person is a complex derived by combining several more basic features, it is im-

possible to have the same node [speaker] marked by + and – simultaneously. But this is just what is needed if fourth person is (or at least can) consist of a combination of first and second person. Therefore, we are forced to allow no negative settings of the nodes on the tree.

In this regard, let us briefly review the feature specifications of each person in turn.

That for first person is thus [+here, +speaker]; second person, by contrast, is [+here, +hearer]; third person singular is distinguished by the feature [+there], whereas third person plural is designated [+there, +plural].

The fact that the Number node [plural] is located only under the [+there] specification encodes the fact that fourth person is not the plural of first person and that fifth person is not always the plural of second person. Fourth person is not, in fact, a plurality of speakers, but includes the speaker and either the hearer and/or somebody else. Fifth person, on the other hand, is sometimes a plurality of hearers, but it can also include somebody who is not present.[18] This is essentially the reason why the terms fourth and fifth person are invoked in this analysis.

Hence, not surprisingly, the feature specification of fourth and fifth persons are more complex and include the activation of several nodes simultaneously. To give a concrete example, fourth person can have different nodes activated, which correspond to the following readings outlined in (16a–f) below:[19]

(16) a. the speaker and only one hearer
      b. the speaker and more than one hearer
      c. the speaker, one hearer and somebody else who is not present
      d. the speaker, one hearer and several persons who are not present
      e. the speaker, more than one hearer and somebody else who is not present
      f. the speaker, more than one hearer and several persons who are not present

This means that in the case of (16a), the feature specification must activate both nodes under [+here], while in all the other cases it must activate the [speaker] node and at least one node which is included under the [+there] specification. Hence, the type of tree illustrated in Figure 4 does not have the formal properties that syntactic trees are standardly assumed to display. It is generally assumed that a syntactic tree has a head whose features are projected up to the maximal projection, while the features of the complement are not. Recently, there have been attempts to derive this fact from more basic properties of syn-

tactic structure, but it has never been called into question since the period when X'-theory was originally formulated and, indeed, it can be traced back at least to Chomsky's (1970) 'Remarks on nominalization' paper. If we intend to follow the idea that Person is a complex entity and that it can be decomposed into more basic features, we have to account for fourth person as a composition consisting of the simultaneous activation of several nodes. Under interpretation (16a) of fourth person these nodes are [+speaker] and [+hearer], hence, the features of both nodes percolate up the tree to the maximal node Person. Other possible feature compositions that one might associate with fourth person are even more complex, because they include feature specifications that belong to two different branches, the [+here] and [+there] components. Once again this relies on the simultaneous activation of two distinct nodes and of nodes dominated by them. The same problem arises with fifth person, which combines the features of second person, namely [+hearer] and those of third person, namely [+there] and [+plural]. The node Person seems thus not to obey the standard restrictions on syntactic structure.

Given that a representation of these persons in a bi-dimensional tree does not seem to satisfy the formal properties that we usually associate with a syntactic tree, we will tentatively propose a different representation, assuming, instead, that the feature activation for these can be captured more faithfully by a structure located in a three-dimensional space, in a similar manner to the hypothesis that basic features of phonemes are located on different planes in autosegmental phonology. According to this theory, phonemes are not simply bundles of unordered features, but correspond to the unification of these on the temporal axis containing only time units (the skeleton), while the features are located on different planes (one plane for nasality, one for the labiality or voice etc.). Several features coming from different planes are connected to the temporal axis forming the phoneme. In this way it is also possible to dissociate the temporal axis from the singles features: geminates thus correspond to one feature bundle connected to two time units on the skeleton which have the same feature composition, while affricates correspond to a single time unit to which the feature bundles of both an occlusive and a fricative consonant are linked.[20] We conceive of the composition of the node Person in a similar way as the simultaneous spell-out of several different features which are located on different planes (see Figure 4). As far as the composition of simpler persons, such as first, second and third, is concerned, it makes no substantive difference whether one assumes a bi-dimensional or a three-dimensional structure. The reason why we think that a three-dimensional structure is preferable, relates

directly to the feature composition of fifth person and to some interpretations of fourth person (essentially all those in (16) except (16a)).

In order to demonstrate our hypothesis more clearly, it will be helpful to review the readings (given in 17a–e below) that can be associated with fifth person:

(17) a.   more than one hearer
      b.   one hearer and somebody else who is not present
      c.   one hearer and several persons who are not present
      d.   more than one hearer and somebody else who is not present
      e.   more than one hearer and several persons who are not present

With the exception of (17a), all the readings listed above are equivalent to those previously associated with fourth person in (16), the only difference being that the [+speaker] feature is not activated.

Already, the reading in (17a) is problematic if we restrict our analysis to the adoption of a bi-dimensional tree. Thus, the features activated by this reading are [hearer] and [plural]. Note however, that [plural] is located under the node [+there]. Hence, the activation of the node [plural] in a bi-dimensional plane automatically also gives us the activation of the [+there] feature, as it is located in the path from the node 'plural' to the node Person. This problem arises under the standard assumption that when a node is activated, its feature percolates up the tree to the nodes that dominate it. Hence, when [Plural] is activated, [+there], the node defining third person, also is. This means, that fifth person should always include a third person, which is clearly not the case. Indeed, it is well known that languages distinguish between inclusive and exclusive person and some account of this important distinction must also be taken in any robust analysis of this issue, even if this distinction is not present in the Romance dialects we know. What we need here, therefore, is a more flexible system of feature composition, than the one the bi-dimensional tree afford, since it forces a unique path which is not always the correct one.[21]

Under the assumption of the three-dimensional tree composed of different planes, which we postulated earlier in this discussion, it would not now be necessary for fifth person to have to percolate up to the [there] node, and the node 'Person' would be confined to the composition of [+hearer] and [+plural].[22] Intuitively, such a formulation captures its feature specifications better than the analysis provided by a bi-dimensional tree. In other words, what is required is a new plane, where only the two relevant features that are activated can be connected to the Person node and the [+there] node remains untriggered as in Figure 4'.[23]
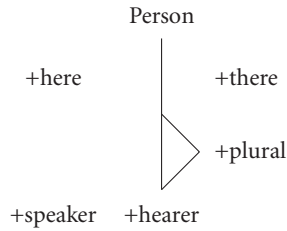
Figure 4'.
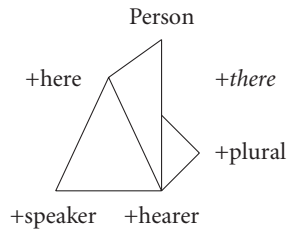
Figure 4".

This third dimension schematized here can also derive more complex cases like those listed in (17b–e), as well as the different interpretations of fourth person listed in (16), and this is obtained by using exactly the same procedure, hence combining the necessary features on a different plane. An additional bonus of this new formulation concerns the fact that the complex feature composition can be obtained, much as in autosegmental phonology, by the combination of different planes. Let us review, for instance, interpretation (16b) of fourth person which, in this model, will have the structure illustrated in Figure 4".[24]

Although first, second and third person do not strictly require a three-dimensional structure, fourth and fifth person do, given that we have to account for the fact that their various interpretations can be morphologically encoded by natural languages and must correspond to the activation of different nodes in the structures given in Figures 4' and 4".

Following this tentative analysis, and in a similar vein to proposals within autosegmental phonology, we should be able to determine how many planes there are and which features they contain. For the moment and with respect to the analysis of persons in Romance dialects, it is possible to hypothesize three planes: one contains the specification of plural, which we have seen to be able to be activated independently from the person it is associated with; the other two

should contain, respectively: (a) the deictic features [+speaker] and [+hearer], and (b) the non-deictic specification from which third person can be derived, namely, the [+**there**] feature alone.

Such a structure cannot be much more than an approximation, at this stage, and needs to be tested on the basis of other language groups. However, our main goal here is not to provide an ultimate characterization of the structure of person features, but to locate the direction in which such a characterization may eventually be sought.

5.   Further evidence for the feature composition

5.1   The peculiarity of fourth and fifth persons

The decisive role of those persons that have a more complex feature composition is twofold. As shown in Section 3 above, fourth and fifth person play a crucial role in the extension patterns of at least two types of elements, namely, reflexive *s+V* forms and vocalic subject clitics in the Northern Italian dialects. Fourth and fifth persons are also significant from a theoretical perspective, since they show that the feature composition that results in the node 'Person' must occur in a three-dimensional space, where different features can be composed without percolating up paths that are obligatory, as in a bi-dimensional tree.

Also of interest from a language universals perspective, is the fact that there is further empirical evidence that fourth and fifth person are 'special', in some sense. Thus, it is noteworthy that their feature composition includes both features belonging to the [+here] domain and features connected to the [+there] domain.

As we feel that these issues are worth exploring further, outlined below is the presentation of some cases which illustrate the particular status of fourth and fifth person both in the verbal and in the pronominal systems, which also has a bearing on our analysis of the feature specification of person.

5.1.1   *Fourth and fifth person in the verb: The present indicative*
Let us point out the formation of the present indicative of irregular verbs that present suppletive forms. It is generally the case that first, second and third persons are derived from one root which is distinct from that of fourth and fifth: the following is the paradigm of the verb *andare* 'go'. First, second and third derive from a root *vad*, fourth and fifth from a base *and*.

(18) a. vado
          "I go"
      b. vai
          "You go"
      c. va
          "He goes"
      d. andiamo
          "We go"
      e. andate
          "You go"
      f. vanno
          "They go"

Other verbs of the same type are all the verbs with the infix -isc- like *finire* 'end', *uscire* 'get out' and *udire* 'hear' (cf. Salvi & Vanelli 2004: 102).[25]

5.1.2 *Fourth and fifth person strong pronouns: Morphological complexity*
Another phenomenon involves strong pronouns: in many dialects of Northern and Southern Italy the adjective *altri* 'others' is added to fourth and fifth person tonic personal pronouns, and not in the third pl.:

(19) a. Ni+altri   Venetian
      b. Vi+altri

While this type of pronoun is extremely common throughout Italy, it is also interesting to note that the forms for possessive pronouns for all persons except fourth and fifth can be shown to be clitics in many of these dialects (as (20) and (21) below from the varieties known as Veneto and Lunigiana demonstrate: see Poletto & Tomaselli 1995 and Penello 2003 for a detailed discussion on this).

Veneto

(20) *clitic*   *free*
      me   'my'   mio/a/i/e
      to   'your'   tuo/a/i/e
      so   'his/her'   suo/a/i/e
      so   'their'   suo/a/i/e
      0   'our'   nostro
      0   'your+pl'   vostro

Some varieties also have reduced forms for the fourth and fifth person *viz.* [nosa] 'our' (Dolomiti) and [vos] 'your+plur' (Bergamasco), but this is un-

common. The Lunigiana dialects show three different series, i.e. (a) clitic
adjective, (b) free adjective and (c) pronoun in predicative position:

(21)   *clitic free predicative*
     mi   me   mei/i/a/me 'my'
     to   tu   tug/i/a/tug 'your'
     so   su   sug/i/a/sug 'his/their'

Of particular significance is the fact that this three-way partition is once again
not found with the fourth and fifth person.

5.1.3 *Fourth and fifth person possessives*

Likewise, other dialects display the same 'resistence' to the creation of fourth
and fifth person clitic forms. Hence, in the Southern Italian domain there are
several dialects that have developed enclitic forms for possessive pronouns with
relational nouns.

Enclitic possessives can be considered a third series, different both from
free elements and from the proclitic forms found in the Northern varieties:
'par<u>du</u> 'your father' (Servigliano Marche) 'patr. <u>m</u>. 'my father' (Abruzzese)
'moye<u>ma</u> 'my wife' (Subiaco, Roma).[26]

There is an implicational scale in the enclitic forms that can be found in a
given dialect: first and second person enclitics are the most frequent. The en-
clitic forms are less common with plural nouns, though they are attested as in,
ne'put<u>imi</u> 'my nephews' (Tagliacozzo, L'Aquila). Third person enclitics are less
frequent than first and second person singular so that in Lucanian, for instance,
only first and second person singular enclitics are attested. The fourth and fifth
person enclitic possessives are the least common of all and can be found only in
a very limited number of dialects: ne'put<u>eno</u> 'our nephew' (Sonnino, Southern
Lazio). Noteworthy in this regard is the fact that this is the same type of dis-
tribution as that found for prenominal possessives in the Northern area. Some
forms even show two enclitics: 'sur<u>tata</u> 'your sister' (Saracena, Calabrian) but,
to the best of our knowledge, this phenomenon is never found with fourth and
fifth person.

We are not alone in these observations, as Penello (2003), for instance,
has also recently noted that very few Romance dialects have enclitic forms for
fourth and fifth person.[27]

We can conclude from this evidence, therefore, that there is a general
tendency in Romance (standard and non-standard) to keep fourth and fifth
person distinct from all other persons. Our explanation in this context is that
this is due to their complex feature composition, which requires the activation

of 'mixed' features of the two domains ([+here] and [+there]), thus necessitating the creation of a distinct plane through which the feature composition gives rise to the complex node Person. The complex fourth and fifth person can have reduced clitic forms, but only when simpler persons do and only in a limited subset of dialects. We believe that their resistence to becoming clitics is related to their complex feature composition. If a clitic form has fewer feature specifications, we expect that those persons which are semantically more complex undergo the process of simplification more slowly; in other words, there is more material to erode.

5.2 Fourth person and impersonal: French *on* and Florentine *si*

One more interesting case that we would like to analyse in this respect, which can arguably reveal something about the mechanisms of extension patterns, is the evidence provided by the extension of the impersonal form to the fourth person in both spoken French and Florentine, exemplified in (22a/b) below:[28]

(22) a. On va   spoken French
          One goes/We go
      b. Si va   Florentine
          One goes/We go

This connection between the form that the impersonal and the fourth person takes in these varieties can only be achieved in a system like the one we have proposed here. In other words, the impersonal form is the one that can contain all the possible feature specifications for all persons; therefore, it is, in a sense, the pronominal that can include all persons since all the feature specifications associated with these are feasible. We will see later on how an extension pattern from the impersonal, as the default case, to the fourth person, as the most complex one, can be captured within our framework.[29]

6.   On deriving extension patterns

In this section we examine the central steps of the extension patterns outlined in Figure 2 and derive them from the three-dimensional structure proposed in Figure 4. Before examining each case we would like to briefly outline a comparison between what we are doing here and what has been done by autosegmental phonology. We have already argued that decomposing the morphology and semantics of the category 'person' is similar to establishing the phonetic features

of a given phoneme. The inventory of phonetic features is universal, but all languages select only some of them and among those only a subset encodes meaning distinctions.

Suppose that among the universal inventory of features Romance languages select those that have been discussed above: so, while they are sensitive to speaker or hearer, they are not sensitive to dual or trial, which can be found in other languages but not within the Romance domain. Inside the set of features that build up the three-dimensional structure illustrated in Figure 4 each Romance language selects a subset which become distinctive, namely, which are reflected into morphological distinctions in the paradigm of pronouns. Each step in the extension pattern illustrated in Figure 2 can be conceived as the deactivation of a single feature, which becomes non-distinctive, thus obliterating the morphological distinction between two different forms and giving the result that one form is extended from one person to the other.

The extension step from fourth to fifth person can be expressed in our structure by assuming that the feature [+speaker] is deactivated, hence it is not more considered to be distinctive. If [+speaker] is not distinctive anymore, the difference between fourth and fifth person is no more encoded in morphology, and the two persons have the same form (in the figure italics and bold represents the deactivation of [+speaker] as a distinctive feature).

The extension step from fifth person to second person also consists in the deactivation of a single feature, namely [plural], hence starting from Figure 5 we obtain Figure 6.

Also the extension from second to first person can be derived in the same way by simply deactivating the feature [+hearer] and leaving only the feature [+here] as distinctive (Figure 7).

Note that it would be impossible to have a different path in the deactivation of the features, for example, deactivating the feature [+hearer] while leaving the feature [+plural] active. This is so because the feature plural is connected to the

Figure 5.

Person

+here   +there

+*plural*

+***speaker***   +hearer

Figure 6.

Person

+here   +there

+*plural*

+***speaker***   +*hearer*
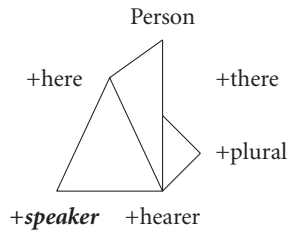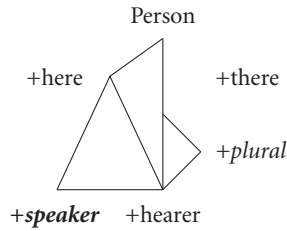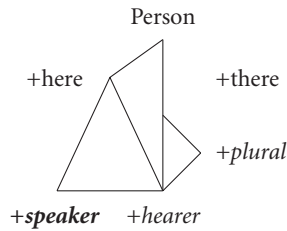
Figure 7.

maximal node Person only through the creation of a plane with [+hearer]. If [+hearer] is suppressed, [+plural] must be suppressed as well, as it does not have any connection to the maximal node.

Note that this also derives the combination of [+speaker] and [+plural] is impossible. This is a welcome result, because, from the semantic point of view there cannot be more than one entity perceived as 'the conscience speaking'.

The last point left to explain is the other side of the implicational scale in Figure 2, namely the reason why third person, more precisely, its impersonal value – and fourth person are connected. In this case the extension cannot simply be obtained by deactivating one feature, but by a different procedure, inverting the activation value of all the features involved: as we hypothesized above, the impersonal reading of third person corresponds to the unmarked value for all features, in other words, the node Person does not contain any feature specification at all (Figure 8).

This is just the opposite of fourth person, which can correspond, at least in one of its readings, to the opposite of an impersonal, namely to the activation of all feature specifications (hence of all nodes dominated by person). Hence, we have to hypothesize that, in addition to the mechanism of deactivation of a single feature, there must be an operation of 'inversion of polarity' of the values of all features and this accounts for the cases of *on* or *si* denoting fourth
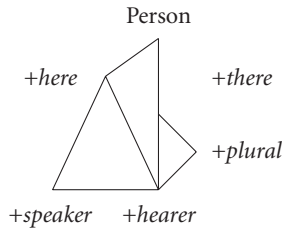
Figure 8.

person discussed in Section 5. We leave further refinements of this operation to future research.

The last point we mention has to do with the relation between our structure, which is based on extension patterns of pronominal forms, and the corresponding structure for DP and adjectival agreement (which look essentially the same). We think that only the system of non-deictic persons is connected to the DP-system, while the system of deictic persons is not related to it. This is also one of the reasons why we think that the two pronominal systems, though related, have to be placed on different planes. We will not discuss the system of DP-agreement here, because we do not have a solid enough empirical basis in our data base yet in order to apply a dialectological perspective as the one we have systematically pursued for pronominal clitics.

7. Conclusion

In this work we have presented and analysed some phenomena concerning the pronoun system of Romance languages, concentrating in particular on Italian dialects and the paths that can be seen as extensions of the value of a given pronoun, which acquires the value of another one and substitutes it. We have shown that fourth and fifth person play an important role in morphological extension patterns. In these processes, fourth and fifth persons appear to act as a 'bridge' for the transition from the deictic to the non-deictic persons. We believe that this is so because the extension process is not a random phenomenon but is determined by the feature composition of the elements that undergo extension, in the sense that extension is limited to cases where the two forms have at least one feature in common. Moreover, extension of this sort is a probabilistic phenomenon: the more features which two forms have in common, the more probable extension there will be. As fourth and fifth persons have

the most complex feature composition, they are very often a nodal point in extension patterns.

On this basis, we have tried to formalize a suitable structure for person features and noted that if we adopt a bi-dimensional tree, the outcome does not have the typical properties of syntactic trees. Therefore, we tentatively propose a three-dimensional structure where features can be combined on more planes to reach the final node Person. This obviously does not take place within the syntactic component, though it remains to be seen exactly what the links are between it and the morphological structure.

Notes

* We thank the editors of this volume for inviting us to take part in this exciting enterprise and for being patient during the various stages of evolution of this chapter, and the anonymous reviewers who helped us to improve and clarify our work. We are particularly grateful to Marc van Ostendorp who provide insightful and detailed comments, which will constitute the input for our future work. For the concerns of the Italian Academy, Paola Benincà takes responsibility over Sections 1 and 2, Cecilia Poletto for Sections 3 to 7. We are aware of the fact that the map found in the appendix is only a rough representation of the dialectal situation in Italy, but we think that it can nonetheless provide the reader with an idea of the variety and complexity of the area we investigate in this work.

1. For theoretical reasons which will become clear during the discussion, we will refer to first person plural as fourth person and to second person plural as fifth person. In referring to 'third person' we intend both singular and plural, if not specified otherwise.

2. One might wonder what a sufficient amount of dialects might be. This is a far-reaching question, which we leave aside here. The number of dialects taken into account here is about 200 in the area of Northern Italy.

3. This is currently being done for the Germanic dialects of the Netherlands by the SAND-project, see http://www.meertens.knaw.nl/projecten/sand/sandeng.html/

4. Romance languages have clitic and tonic pronouns. Clitic pronouns have a constrained distribution, as they cannot be modified, coordinated, used in isolation and focalized. Moreover, they occur in a fixed position in the clause (in most Romance languages attached to the verb). The Northern Italian dialects also have subject clitics of at least four different types, and some of them can co-occur; Therefore, several examples have two sets of subject clitics, both glossed as SCL. A subset of them are vocalic clitics, which have special syntactic properties.

5. We use the abbreviation s+V for the forms *si/se/sa/* found in various dialects.

6. Kayne (2002) hypothesises that the fourth and fifth person *n+V* and *v+V* in French are in the same paradigm with *m* and *t*. It is important to note that this is not true of Italian *ci* and *vi*, whose origin is from a locative pronoun.

7. Notice that for some persons a subject clitic also occurs, while for others it does not. This is tangential to our present discussion and we refer to Benincà (1994) and Poletto (2000) for a detailed analyses of subject clitics (see Note 13).

8. The AIS: 'Atlas of Italy and Southern Switzerland' is a dialect Atlas in seven volumes; it contains mainly lexical and phonetic data, but also verbal morphology presented in a rather systematic way. Syntactic data are not systematically displayed but occur in several maps.

9. In the declarative sentence a subject clitic is present, whereas in the interrogative clause there is, instead, an enclitic due to the process of subject clitic inversion, which is mandatory in main interrogatives.

10. We refrain here from giving a theoretical analysis of this phenomenon, although it is probably connected to the presence of a set of projections for clitics also contained within the CP-layer (as proposed, among others, by Uriagereka 1995 and Sportiche 1996) and that these can be activated only when the verb itself moves into the CP domain.

11. This example is a case of optional doubling of the reflexive, which is also found with second and first person in other dialects.

12. The only exception that we have found to our pattern is the dialect spoken in Trieste, where *se* has extended to the second person but not to the fifth. We do not know how to explain this case, but note that this is a dialect on the border with Slovenia and the pattern obtaining here, therefore, may be due to language contact.

(i) a. El   se   ga meso i   calzeti   *Trieste*
         he himself has put the socks
         "He has put on his socks"
     b. I   se   ga   meso i   calzeti
         they themselves have put the socks
         "They have put on their socks"
     c. Se   gavemo meso i   calzeti
         ourselves have   put the socks
         "We have put on our socks"
     d. Te se   ga   meso i   calzeti
         you yourself have put the socks
         "You have put on your socks"

13. These have special syntactic properties – such as the fact that they interact with typical CP elements – and have also been analysed as belonging to the CP layer (cf. Benincà 1983 and Poletto 2000); this singles them out with respect to the more usual "agreement clitics" located within the IP layer.

14. These examples are taken from a XVI century text, namely *Le Lettere* by Andrea Calmo, written in the dialect (in fact, 'language' from the sociolinguistic point of view) spoken and written at that time in Venice.

15. In (13b) there are two subject clitics, a phenomenon which is quite frequent in the NIDs (see Note 4).

16. See Section 5 for examples on this point.

17. Even in Sardinian, where the determiner comes from the Latin word for the anaphoric pronoun "this" *IPSE,* the pronoun is still a form derived from *ille*.

18. Kayne (2002) assumes that the personal and possessive pronouns of fourth person have a plural morpheme attached. In Italian, this plural morpheme would then exceptionally be an *s*, like it is in French, a morpheme that only marks plural in a few Northern dialects and not in Italian. We propose that the morpheme can be treated as marking complexity of features and not plurality *per se*.

19. M. van Ostendorp has pointed out to us that we do not consider here fourth person with exclusive interpretation (i.e. referring to first person + third person and excluding the hearer). This is because, as far as we know, there is no distinction between 'inclusive' and 'exclusive' fourth person in Romance. The present system is most likely to be the result of simplification of more complex systems which did used to included the distinction 'inclusive' vs. 'exclusive'. Traces of the preceding stage are present in various dialects, in the form of the pronoun, whose value in any case is ambiguous. We hope to have the opportunity to deal with these more complex systems in future work. Moreover, our hypothesis predicts that there are extensions between third person and fourth person, as the only difference between them concerns the [+speaker] feature. This is indeed confirmed, for instance, by the case of extension of the reflexive clitic form from the third person to the fourth person.

20. The reader should be aware that the similarity with autosegmental phonology and our system is offered as an entirely impressionistic suggestion. All phonological features are activated on separate planes and, while here the present research does not examine the possibility of having separate planes, but rather attempts to building an object that is more similar to a neural net, with different paths inside a complex three-dimensional web. Likewise, in autosegmental phonology three-dimensionality is not immediately apparent in trees but in specific processes, at least this is the contention of some phonologists (see for example Halle 1995; Calabrese 1995). The spreading of features can involve non-adjacent segments if the targets are adjacent in the relevant plane, so long as within the three-dimensional space the intervening segment has no specifications in that plane. It is also important to note in this regard that the connection between 'Place' node and 'place' features are better conceived of as being able in a three-dimensional space.

21. There is another logical possibility that we discard here, namely that the [plural] feature is located both under the [+there] node and under the [+hearer] node. This seems to us to be an unnecessary duplication that cannot be justified. An anonymous reviewer noted that reduplication of features is very common in natural languages. We are aware of this fact, but we think that redundancy is a radically different case from the one we are dealing with here: assuming a reduplication of features would introduce doubling inside the theory, not at the empirical level. Although the empirical level displays a complex and redundant set of data, our theory must be as simple and minimal as possible. Reduplication of features, in other words, is observed as a multiple expression of the same grammatical fact (for example, in many agreement facts) but not inside the same projection as a multiple occurrence of the same feature, not in phonology nor in syntax or morphology. Therefore, we will not admit reduplication of features, although redundancy exists.

22. The [+plural] feature is thus testricted to being connected either to the [+there] feature or to the [+hearer] feature, and not to the [+speaker] feature. This conclusion reflects the

assumption, which we take to be a 'primitive', namely, that the speaker is cognitively 'unique', so that there cannot properly be a plurality of speakers. To obtain, from a different perspective, see Giorgi and Pianesi (2005), who also presents a number of empirical arguments in favour of this assumption.

23. This structure does not include any gender feature, because in the two extension patterns gender never appears to play any role. Nevertheless, given that this hypothesis intends to cover also other possible, if not yet known, cases of pronominal extension, gender should also be taken into account. Although we do not insert it into our figures for the above reasons, it is fairly easy to see that gender should be under the [+there] feature in Romance, because it only occurs in the system of third persons, and not in the system of deictic persons. We leave this point for future research, when extension patterns including pronominal forms with gender will have been explored, if they exist.

24. The fact that the feature [+there] is not activated by the configuration of fourth person is expressed here by putting into italics the inactive feature. Italics means deactivation of the feature in the other figures as well.

25. In some cases there might be interferences from phonological constrainsts, as the fourth and the fifth persons are the only ones where the accent is not on the stem.

26. Hence, postnominal, prenominal and enclitic possessives could be considered to be three different classes. This could lead to our considering the dichotomy clitic/free from a different perspective, in the sense that the splitting into more than one series is a complex phenomenon regarding the internal structure of pronominal phrases (for a detailed theoretical discussion see Cardinaletti & Starke 1999).

27. The special status of fourth and fifth person is also signalled by the phenomenon of partial drop in Middle French. It is well known (cf. Roberts 1993 inter alia) that in Middle French, the only persons that have *pro*-drop in all contexts (main and embedded declaratives, interrogatives etc.) are precisely fourth and fifth person.

28. The same pattern is attested in other dialects such as some varieties of the Bergamo area, where the form originally deriving from Latin HOMO is extended from an impersonal to the fourth person, and Friulian, where impersonal *si* is used both for impersonal and fourth person.

29. We are very well aware that one meets serious problems when trying to connect in a formal way the two stages first plural → impersonal by some formal device. Marc von Ostendorp (p.c.) has suggested some interesting possibilities to obviate these solutions, which we will hopefully take into consideration in the future. Here we merely note that the extension process we have suggested in all probability follows the exactly the path we have indicated in this paper.

References

AIS, Sprach- und Sachatlas Italiens und der Südschweiz (1928–1940). Rengien, Zofingen.
Benincà, Paola (1983). "Il Clitico 'a' nel Dialetto Padovano." In *Scritti Linguistici in Onore di Giovan Battista Pellegrini* (pp. 25–32). Pisa: Pacini. Reprinted in Benincà (1994).

Benincà, Paola (1994). *La Variazione Sintattica.* Bologna: Il Mulino.

Benincà, Paola & Laura Vanelli (1982). "Appunti di Sintassi Veneta". In M. Cortelazzo (Ed.), *Guida ai Dialetti Veneti,* Vol. 4 (pp. 7–38). Padova: CLEUP. Reprinted in Benincà (1994).

Benveniste, Emile (1966). "Structure des Relations de Personne dans le Verbe". In *Problèmes de Linguistique Générale* (pp. 195–204). Paris: Gallimard.

Bianchi, Valentina (2004). "On Finiteness as Logophoric Anchoring". In J. Guéron & L. Tasmovski (Eds.), *Temps et Point de Vue / Time and Point of View* (pp. 213–246). Université de Paris X Nanterre.

Calabrese, Andrea (1995). "A constraint based theory of phonological markedness and simplification procedures". *Linguistic Inquiry, 26,* 373–463.

Calmo, Andrea (1888). *Le Lettere.* V. Rossi (Ed.). Torino: Loescher.

Cardinaletti, Anna & Michal Starke (1999). "The Typology of Structural Deficiency: On the Three Grammatical Classes". In H. van Riemsdijk (Ed.), *Clitics in the Languages of Europe. Empirical Approaches to Language Typology*, Vol. 8 (pp. 145–233). Berlin: Mouton de Gruyter.

Cornips, Leonie & Cecilia Poletto (2005). "On Standardising Syntactic Elicitation Techniques. PART I". *Lingua, 115*(7), 939–957.

Giorgi, Alessandra & Fabio Pianesi (2005). "Credo (I believe): Epistemicity and the syntactic representation of the speaker". Ms. University of Venice.

Halle, Morris (1995). "Feature geometry and feature spreading". *Linguistic Inquiry, 26,* 1–46.

Harley, Heidi & Elizabeth Ritter (2002). "Person and Number in Pronouns: A Feature-Geometric Analysis". *Language, 78*(3), 482–526.

Kayne, Richard (2002) *Parameters and Universals.* New York: Oxford University Press.

Penello, Nicoletta (2003). Capitoli di Morfologia e Sintassi del Dialetto di Carmignano di Brenta. Ph.D. Thesis. University of Padova.

Poletto, Cecilia (2000). *The Higher Functional Field*: *Evidence from Northern Italian Dialects.* New York: Oxford University Press.

Poletto, Cecilia & Alessandra Tomaselli (1995). "Verso una Definizione di Elemento Clitico". In R. Dolci & G. Giusti (Eds.), *Quaderni di Linguistica del CLI* (pp. 159–224). Venezia: CLI.

Roberts, Ian (1993). *Verbs and Diachronic Syntax: A Comparative History of English and French*. Dordrecht: Kluwer.

Salvi, Giampaolo & Laura Vanelli (2004). *Nuova Grammatica Italiana.* Bologna: Il Mulino.

Shlonksy, Ur & Isabelle De Crousaz (2003). "The Distribution of a Subject Clitic Pronoun in a Franco-Provençal Dialect". *Linguistic Inquiry, 34*(3), 413–442.

Sigurdsson, Halldor (2004). "The Syntax of Person, Tense and Speech Features." Ms., University of Lund.

Uriagereka, Juan (1995). "Aspects in the Syntax of Clitic Placement in Western Romance". *Linguistic Inquiry, 26,* 79–123.

Vassere, Stefano (1993). *I Pronomi Clitici del Luganese.* Milano: Franco Angeli.

Appendix

Map of Italian dialects

Aree dialettali

▮	Provenzale	▮	Meridionale-Interno
▮	Franco-Provenzale	▮	Meridionale-Estremo
▮	Gallo-Italico	▮	Sardo-Logudorese
▮	Veneto	▮	Sardo-Campidanese
▮	Ladino	▮	Sardo-Sassarese
▮	Tirolese	▮	Sardo-Gallurese
▮	Friulano	▒	Aree miste di complessa classificazione
▮	Sloveno	– – – –	Limite approssimativo di varietà dialettali
▮	Toscano	———	Decorso di isoglosse
▮	Mediano		

Isole alloglotte

Ⓐ	Albanese	Ⓔ	Emiliano
Ⓖ	Greco	Ⓒ	Catalano
Ⓟ	Provenzale	Ⓣ	Tedesco
Ⓗ	Croato	Ⓢ	Gallo-Italico (Settentrionale)
Ⓛ	Ligure	Ⓕ	Franco-Provenzale

Isoglosse

A   Conservazione dei nessi con l

B   Palatalizzazione di ca e ga

C   Evoluzione di a in è

D   Evoluzione di ù in ü

E   Lenizione delle consonanti sorde

E1   Lenizione di p in v

E2   Lenizione di c in g

E3   Lenizione di t in d

E4   Degeminazione delle consonanti

F   Assimilazione di nd a nn e di mb a mm

G   Distinzione tra –u ed –o

H   Passaggio delle vocali finali alla vocale indistinta ə

I   Sonorizzazione delle sorde dopo n

L   Posposizione del pronome possessivo

M   Riduzione di pl e cl a kj e č

N   Limite dell'area "Lausberg" (vocalismo arcaico)

O   Limite settentrionale del vocalismo "siciliano" in Calabria e nel Salento

P   Limite meridionale della riduzione della vocale finale ə

Q   Limite settentrionale della fusione di –e ed –i in i nel Salento

R   Limite meridionale del vocalismo con metafonesi

S   Limite settentrionale della conservazione dei nessi nd ed mb

T   Limite settentrionale della mancanza del passato prossimo

U   Limite settentrionale della mancanza dell'infinito

Z   Limite dell'area siciliana nord-orientale un tempo prevalentemente grecofona

Isofone Tirolesi

A   Diminutivo dei plurisillabi in ll

B   Fonetica: ě in e

C   Fonetica: uo in ui

DIALETTI
Ⓐ

Index

CURRENT ISSUES IN LINGUISTIC THEORY

E. F. K. Koerner, Editor

Zentrum für Allgemeine Sprachwissenschaft, Typologie
und Universalienforschung, Berlin
efk.koerner@rz.hu-berlin.de

Current Issues in Linguistic Theory (CILT) is a theory-oriented series which welcomes contributions from scholars who have significant proposals to make towards the advancement of our understanding of language, its structure, functioning and development. CILT has been established in order to provide a forum for the presentation and discussion of linguistic opinions of scholars who do not necessarily accept the prevailing mode of thought in linguistic science. It offers an outlet for meaningful contributions to the current linguistic debate, and furnishes the diversity of opinion which a healthy discipline must have. A complete list of titles in this series can be found on the publishers website, **www.benjamins.com**

248 **WEIGAND, Edda (ed.):** Emotion in Dialogic Interaction. Advances in the complex. 2004. xii, 284 pp.

247 **PARKINSON, Dilworth B. and Samira FARWANEH (eds.):** Perspectives on Arabic Linguistics XV. Papers from the Fifteenth Annual Symposium on Arabic Linguistics, Salt Lake City 2001. 2003. x, 214 pp.

246 **HOLISKY, Dee Ann and Kevin TUITE (eds.):** Current Trends in Caucasian, East European and Inner Asian Linguistics. Papers in honor of Howard I. Aronson. 2003. xxviii, 426 pp.

245 **QUER, Josep, Jan SCHROTEN, Mauro SCORRETTI, Petra SLEEMAN and Els VERHEUGD (eds.):** Romance Languages and Linguistic Theory 2001. Selected papers from 'Going Romance', Amsterdam, 6–8 December 2001. 2003. viii, 355 pp.

244 **PÉREZ-LEROUX, Ana Teresa and Yves ROBERGE (eds.):** Romance Linguistics. Theory and Acquisition. Selected papers from the 32nd Linguistic Symposium on Romance Languages (LSRL), Toronto, April 2002. 2003. viii, 388 pp.

243 **CUYCKENS, Hubert, Thomas BERG, René DIRVEN and Klaus-Uwe PANTHER (eds.):** Motivation in Language. Studies in honor of Günter Radden. 2003. xxvi, 403 pp.

242 **SEUREN, Pieter A.M. and Gerard KEMPEN (eds.):** Verb Constructions in German and Dutch. 2003. vi, 316 pp.

241 **LECARME, Jacqueline (ed.):** Research in Afroasiatic Grammar II. Selected papers from the Fifth Conference on Afroasiatic Languages, Paris, 2000. 2003. viii, 550 pp.

240 **JANSE, Mark and Sijmen TOL (eds.):** Language Death and Language Maintenance. Theoretical, practical and descriptive approaches. With the assistance of Vincent Hendriks. 2003. xviii, 244 pp.

239 **ANDERSEN, Henning (ed.):** Language Contacts in Prehistory. Studies in Stratigraphy. Papers from the Workshop on Linguistic Stratigraphy and Prehistory at the Fifteenth International Conference on Historical Linguistics, Melbourne, 17 August 2001. 2003. viii, 292 pp.

238 **NÚÑEZ-CEDEÑO, Rafael, Luis LÓPEZ and Richard CAMERON (eds.):** A Romance Perspective on Language Knowledge and Use. Selected papers from the 31st Linguistic Symposium on Romance Languages (LSRL), Chicago, 19–22 April 2001. 2003. xvi, 386 pp.

237 **BLAKE, Barry J. and Kate BURRIDGE (eds.):** Historical Linguistics 2001. Selected papers from the 15th International Conference on Historical Linguistics, Melbourne, 13–17 August 2001. Editorial Assistant: Jo Taylor. 2003. x, 444 pp.

236 **SIMON-VANDENBERGEN, Anne-Marie, Miriam TAVERNIERS and Louise J. RAVELLI (eds.):** Grammatical Metaphor. Views from systemic functional linguistics. 2003. vi, 453 pp.

235 **LINN, Andrew R. and Nicola McLELLAND (eds.):** Standardization. Studies from the Germanic languages. 2002. xii, 258 pp.

234 **WEIJER, Jeroen van de, Vincent J. van HEUVEN and Harry van der HULST (eds.):** The Phonological Spectrum. Volume II: Suprasegmental structure. 2003. x, 264 pp.

233 **WEIJER, Jeroen van de, Vincent J. van HEUVEN and Harry van der HULST (eds.):** The Phonological Spectrum. Volume I: Segmental structure. 2003. x, 308 pp.

232 **BEYSSADE, Claire, Reineke BOK-BENNEMA, Frank DRIJKONINGEN and Paola MONACHESI (eds.):** Romance Languages and Linguistic Theory 2000. Selected papers from 'Going Romance' 2000, Utrecht, 30 November–2 December. 2002. viii, 354 pp.

231 **CRAVENS, Thomas D.:** Comparative Historical Dialectology. Italo-Romance clues to Ibero-Romance sound change. 2002. xii, 163 pp.

230 **PARKINSON, Dilworth B. and Elabbas BENMAMOUN (eds.):** Perspectives on Arabic Linguistics. Papers from the Annual Symposium on Arabic Linguistics. Volume XIII-XIV: Stanford, 1999 and Berkeley, California 2000. 2002. xiv, 250 pp.

229 **NEVIN, Bruce E. and Stephen B. JOHNSON (eds.):** The Legacy of Zellig Harris. Language and information into the 21st century. Volume 2: Mathematics and computability of language. 2002. xx, 312 pp.

228 **NEVIN, Bruce E. (ed.):** The Legacy of Zellig Harris. Language and information into the 21st century. Volume 1: Philosophy of science, syntax and semantics. 2002. xxxvi, 323 pp.

227 **FAVA, Elisabetta (ed.):** Clinical Linguistics. Theory and applications in speech pathology and therapy. 2002. xxiv, 353 pp.

226 **LEVIN, Saul:** Semitic and Indo-European. Volume II: Comparative morphology, syntax and phonetics. 2002. xviii, 592 pp.

225 **SHAHIN, Kimary N.:** Postvelar Harmony. 2003. viii, 344 pp.

224 **FANEGO, Teresa, Belén MÉNDEZ-NAYA and Elena SEOANE (eds.):** Sounds, Words, Texts and Change. Selected papers from 11 ICEHL, Santiago de Compostela, 7–11 September 2000. Volume 2. 2002. x, 310 pp.