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. The contributions in Part III by Adger and Smith, King and Van Gelderen seek to address syntactic variability in real and appa-
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. 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);

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. 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?

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 (1998) and Foulaikes et al. (1999) which examine variable caretaker input and child output, although their investigations focus solely on the phonological component. Extending this research “above and beyond phonology” (Sanlof 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 polylectic (cf. Auer ms. and Harris 1985). Indeed, as Henry, this volume demonstrates more attention should be paid by both sociolinguists and bilingualists 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 Acadia 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 passion 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 of focus-raising across divergent Hungarian lects); (ii) dialects (like Barbiers’ account of word order strategies in Dutch varieties and Benincà and 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 elicitation 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 topic of parameter setting/re-setting as well as pro-drop and verb second phenomena \textit{inter alia}. The unusual patterning of verbal agreement, \textit{do}-periphrasis and negation in Yucatec English is tackled by Adger and Smith while Cheshire’s paper revisits conventional sociolinguistic accounts of variation in pronoun tagging and verbal \textit{-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 \textit{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 Bobaljik & Jonas 1996; Pollock 1989; Rizzi 1997 \textit{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 Wurf 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 1999: Van der Wurf 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, dialectically 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."

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; Eskinovits 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. Herefore, spontaneous data within the variationist paradigm has been considered to be far more authentic than in-

---

5. henry's contributions are discussed in more detail in the following section.
tuitions or elicited data of the sort advocated in some of the contributions to this volume (cf. Coupland 2003 and Rickford 1987). 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 (Barbier 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 pre-fabricated 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 headness 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.

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.

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 (optionalty) 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 historical 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) Ideolectal or personal pattern variation which is the focus of Henry's and Gervain and Zemplèn'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 Zemplèn, 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 substratal 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 Ecuadorian 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 (2003:93f), in a sociolinguistic context, has recently termed "Oddballs" and "Insiders", i.e. individualities 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 adverbal 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-sociolinguistically 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 §§5.1–3.2 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 focused 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 Kyö & Waller 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 1990s (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 Mynskes, 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 Barbier 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 auxiliary 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 vocative 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 fervor 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

The editors would like to close by acknowledging the financial support provided to various phases of this project by the School of English Literature, Language and Linguistics, University of Newcastle, The Meeters Institute and the British Council/Netherlands Organization for Scientific Research (grant no: JRP643).

We would also like to express our deeply felt gratitude to our authors who have gracefully endured our cajoling and actively engaged with us in finding a bridge that connects two disciplines that have stood apart for far too long. There are also a number of other people who deserve special thanks including: Konrad Koerner, the series editor responsible for this volume, for his helpful feedback; Anke de Looper, acquisition editor of the series, for her patience with our many technical queries; Aniek Mennings and Claire Stevenson for their assistance with formatting and indexing; the organising and scientific committees of the Second International Conference on Language Variation and Change and the following colleagues who generously gave of their time and submitted various chapters of the volume to critical, stylistic and formal scrutiny: Peter Auer, Hans Bennis, Hans Bruckhuis, Norbert Corver, Ineke Van de Craats, Liliane Haegeman, Anders Holmberg, Jim Miller, Jim Milroy, Herman Mois, John Nerbonne, Susan Pintzuk, Geoff Poole, Shana Poplack, Ian Roberts, Sali Tagliamonte, Tonjes Veenstra and Martha Young-Scholten.

Notes

2. 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.

3. 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.

4. Sociolinguistic models have, in fact, had considerable success in unravelling 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 definitives 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, semantic 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 relevance 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.
References


