Why grammar is real: a usage-based perspective on patterns

Lieselotte Anderwald and Benedikt Szmrecsanyi

1. Introduction

This paper seeks to explore the grammar-versus-lexis issue in a variationist perspective. Our point of departure is the fact that most research into collocations, colligations, multi-word units, constructions and related phenomena has taken observable patterns – and we acknowledge that naturalistic speech *is* highly patterned – as given. Yet it is not clear why, in a particular instance, the analyst sees some specific pattern A instead of some equally acceptable alternative pattern B. In this connection, we are interested in the extent to which lexical co-occurrence patterns can help us to explain synchronic linguistic variation. We note along these lines that decades of research into grammaticalization have shown that *diachronically*, grammatical structure derives from lexis – but the question is how explanatory lexical patterns are for variation in *synchrony*. In conclusion, we shall argue that rumors of the death of grammar are much exaggerated.

Note now that the term GRAMMAR is used in widely different senses by different scholars. We therefore wish to offer a concise definition of what we understand by the term at the outset of this paper. We define grammar as *the set of systematic, abstract generalizations about the organization of linguistic units*. Hence, a grammar details abstract generalizations about linguistic patterns; crucially, an abstract generalization is one that is not tied to specific lexical items. To illustrate: the observation that the lexical item, or adjective, *large* tends to co-occur with lexical items, or nouns, referring to quantities and sizes (Stubbs 2001: 65) is certainly a worthwhile and interesting observation, but *per se* it cannot be generalized to apply to other adjectives. This is why, in our definition, this statement does not qualify as a grammatical statement, or even as an observation touching on grammatical facts. Real generalizations about the organization of linguistic units run along the lines of statements such as 'in English, main clauses have SV word order', 'adjectives co-occur with nouns' or – even more specifically – 'adjectives specifiying sizes co-occur with nouns referring to quantities'.

This paper is structured as follows: In section 2, we endeavor to quantify the explanatory power of lexical patterns for explaining linguistic variation. In section 3, we demonstrate that type frequency, as a comparatively abstract predictor, is strongly implicated in the distribution of irregular verbal forms in non-standard varieties of English. In section 4, we present evidence that markedness plays an important role in non-standard negation paradigms. Section 5 will provide some concluding remarks.

2. On the probabilistic impact of lexical bias in variationist frameworks

This section will demonstrate that lexical preferences can be modeled and quantitified in variationist-probabilisitc frameworks, and that these preferences – while significant – are insufficient, vis–à-vis non-lexical factors, to adequately explain the linguistic variability at hand. We will take this as evidence that abstract grammars are empirically and psychologically real.

Most analysts interested in why speakers and writers use some variant form instead of some other, equally acceptable variant form operate, in some way or another, in a variationist-probabilistic framework. The research question is typically: What are the non-categorical, statistical factors that have a bearing on language users' linguistic choices?

The basic tenets underlying approaches in this spirit can be summarized as follows:

1. Grammar is multivariate in nature. This means that linguistic choices are typically subject to a number of conditioning factors – be they lexical, semantic, discourse-functional, processing-related, and so on (cf. Gries 2005a).
2. These conditioning factors are not necessarily innate (although they may be – for instance, processing-related factors such as the principle of Early Immediate Constituents (EIC) [cf. Hawkins 1994] is likely to be ultimately due to innate properties of the human parser).
3. The effect of conditioning factors is typically statistical and gradient, not categorical and discrete (cf. Bod, Hay, and Jannedy 2003).
4. Hence, a 'grammar' is a multivariate system of statistical regularities.

Variationist-probabilistic approaches to grammatical variation, then, firmly belong to the family of usage-based approaches. The added bonus is that variationist-probabilistic approaches are rigorously quantitative, utilizing advanced statistical analysis techniques such as regression analysis to shed light on the gradient nature of both linguistic knowledge and usage.

To illustrate this line of research, we will review in this section the contribution of lexical bias (more specifically, verb bias) to variability in particle placement, a binary alternation after transitive, separable phrasal verbs where speakers of English are free to choose between either ordering of particle (henceforth: Prt) and direct object (henceforth: DO), as in (1):

(1) a. Mary *looked* the word *up* (V DO Prt)

b. Mary *looked up* the word (V Prt DO)

This alternation advertizes itself as a case study: for one thing, particle placement is comparatively well studied, particularly in regard to verb bias. Gries and Stefanowitsch (2004), for example, show that *carry out* has a marked preference for the V Prt DO pattern, while *get back* has a strong tendency to take the V DO Prt pattern. On the other hand, the alternation is demonstrably susceptible to a number of other, non-lexical conditioning factors (semantic, discourse-related, and processing-

related; cf. Szmrecsanyi 2005, 2006 for an overview), a circumstance that enables the analyst to systematically test the effect of verb bias against the combined effect of other, non-lexical conditioning factors.

In this spirit, we will re-analyze the dataset used in Szmrecsanyi (2005, 2006) to model how, in naturalistic speech, language user's word order choices can be thought of as a function of several conditioning factors, one of which is verb bias. N = 1,168 particle-verb occurrences drawn from the *Freiburg English Dialect Corpus* (FRED) (cf. Szmrecsanyi and Hernández 2007) were annotated for the following language-internal conditioning factors:

- LENGTH OF THE DO (in syllables): according to the principle of end weight, longer DOs should prefer the V Prt DO ordering (cf. Behaghel 1909/1910);
- SYNTACTIC COMPLEXITY OF THE DO: also thanks to end weight, syntactically complex DOs should prefer the non-separated ordering (cf. Behaghel 1909/1910);
- INFORMATION STATUS: discourse-old DOs should prefer the V DO Prt ordering (for instance, Chen 1986);
- PERSISTENCE/SYNTACTIC PRIMING: for psycholinguistic reasons, the syntactic option previously chosen should *per se* be more likely to be chosen again (for instance, Gries 2005b);
- VERB BIAS: particle verbs have distinctive collostruction strengths (Gries and Stefanowitsch 2004).

Since the data source is a dialect corpus sampling different dialects from all over Great Britain, it stands to reason that we also include a language-external variable (DIALECT AREA) in our analysis, even though dialectal differences are not of primary interest in the present study. To pinpoint the quantitative contribution of verb bias to particle placement variability, we utilize *binary logistic regression*, a multivariate statistical analysis technique which is applicable "wherever a choice can be perceived as having been made in the course of linguistic performance" (Sankoff 1998: 151). Logistic regression has the following advantages over monofactorial analysis methods: (i) it seeks to predict a binary outcome (i.e. a linguistic choice) given several independent (or predictor) variables; (ii) it quantifies the influence of each predictor; (iii) it specifies the direction of the effect of each predictor; (iv) it states how much of the empirically observable variability is explained by the predictors considered (the relevant measure we will report is Nagelkerke R^2, a pseudo R^2 measure for logistic regression that we will interpret, for the sake of simplicity, as indicating the share of variability explained); and (v), it states how well the model fares in predicting actual speakers' choices. What will take center stage here is (iv), i.e. the precise share of variability accounted for by the predictor variables included in the model.

We regressed particle placement choices (V DO Prt vs. V Prt DO) against the five language-internal conditioning factors detailed above, in addition to dialect area as a language-external predictor. The resulting model is highly significant (χ^2 = 291.110, df = 8, $p < .001$), accounts for a fair amount of the overall variability in particle

placement in the dataset (Nagelkerke R^2 = .331), and correctly predicts 82.1% of language users' particle placement decisions in the dataset (baseline: 75.9%).

	partial Nagelkerke R^2	significance
DIALECT AREA	.101	$p < .001$
LENGTH OF THE DO	.088	$p < .001$
VERB BIAS	.029	$p < .001$
PERSISTENCE/SYNTACTIC PRIMING	.015	$p < .001$
SYNTACTIC COMPLEXITY	.009	$p = .004$
INFORMATION STATUS	.005	$p = .041$

Table 1: Partial Nagelkerke R^2 values associated with individual conditioning factors

Table 1 reports partial Nagelkerke R^2 scores for each of the individual variables included in the model, indicating how crucial they are, from a bird's eye perspective, for predicting particle placement (more specifically, what is displayed in Table 1 is the decrease in model Nagelkerke R^2 if individual variables are removed from the fully specified model; cf. Hinrichs and Szmrecsanyi 2007: 463-464 for a similar method). The results reveal that geographic variability (predictor DIALECT AREA) accounts for roughly 10% of the overall particle placement variability in the dataset; LENGTH OF THE DO accounts for roughly 9% of the variability; VERB BIAS – as operationalized through collustruction strength scores – accounts for 3%; and so on. In all, Table 1 makes amply clear that while VERB BIAS is a highly significant predictor, it is a good deal less important than other conditioning factors, such as the principle of end weight (operationalized here as the length of the direct object in syllables). The bottom line, then, is that the analyst who would abandon the quest for abstract, non-lexical generalizations in favor of a view that reduces particle placement variability to lexical preferences could only explain a minuscule share of the observable variability.

It is important to note that particle placement is by no means unique in terms of the overall relevance of lexical bias in grammatical variation. Consider the following, additional evidence from other binary alternations in the grammar of English:

- In the case of clausal complementation (e.g. *I think that John is my friend* versus I *think __ John is my friend*) in conversational British English as sampled in ICE-GB (cf. Kolbe and Szmrecsanyi 2008), verb bias (i.e. *to think* versus *to say* versus *to know*) explains less than 1% of the overall variability, and is hugely less explanatory than other conditioning factors, such as the length of the embedded subordinate clause.
- Re-examining the verbal complementation (e.g. *I started wondering* versus *I started to wonder*) dataset utilized in Szmrecsanyi (2006: 153-180) and drawn from the spoken-conversational section of the *British National Corpus*, it emerges that verb bias (*to start* versus *to begin* versus *to help* etc.) accounts for roughly 11% of the overall variability. This share is substantially higher than in the alternations

discussed so far, but even in the case of verbal complementation, other, more abstract predictors have more explanatory power.

We would also like to add a few additional comments about morphosyntactic persistence (also known as *priming*) at this point. Recall that this phenomenon describes language users' tendency to recycle patterns used or heard in previous discourse. As can be seen from Table 1, the phenomenon is responsible for ca. 1.5% of the overall variability in particle placement, and hence has less explanatory potency than verb bias. Still, we submit that the mere existence of the phenomenon plays havoc with claims that grammar – understood here as word order generalizations – is wholly epiphenomenal to lexis. This is because previous research (Gries 2005b; Szmrecsanyi 2005, 2006) indicates that the phenomenon can operate across particle verb occurrences *with different verb lemmas*. So, not only can an occurrence of the V DO Prt pattern with, say, the verb *look up* (as in (1a)) prime the word order pattern of a later occurrence of the verb *look up* (as in (1b)), but it can also prime a later occurrence of a lexically entirely different particle verb (e.g. *bring in*, as in (1c)).

(2) a. Mary *looked* the word *up* (V DO Prt)

b. Tom *looked* the report *up* (V DO Prt)

c. Tom *brought* the Christmas tree *in* (V DO Prt)

This, in our view, is unassailable evidence that language users store and replicate purely positional syntactic information, and that abstract grammatical patterns which are independent from lexis are psychologically real (for a genuinely psycholinguistic account, see Pickering and Branigan 1998).

By way of an interim summary, the discussion in this section has three principal findings. First, the explanatory potency of lexical bias vis-à-vis other conditioning factors is amenable to precise calculation in probabilistic-variationist approaches to grammatical variation. Second, we have seen that verb bias plays a significant but relatively minor role in predicting language users' word order choices, and that other, more abstract conditioning factors – i.e. the principle of end weight – are a good deal more important. Lastly, this section has highlighted the fact that language users appear to store and recycle purely positional patterns, which is evidence that grammar cannot be reduced to purely lexical co-ocurrence patterns.

3. On the role of type frequency in inflectional morphology

In non-standard data, we can observe recurrent patterns in inflectional morphology that are particularly striking. Our previous in-depth studies (Anderwald 2007, 2009) have shown that non-standard verb forms cluster around the standard English paradigm *cling – clung – clung*, and that in particular verbs that have three distinct forms in standard English (*drink – drank – drunk* or *swim – swam – swum*) tend to have just two forms in many non-standard varieties of British English. Here, variation is almost non-existent, and paradigms typically have a different past tense form (namely in <u>), but keep the past participle identical to standard English. In other words, a typical non-standard paradigm would be *drink – drunk – drunk* or *swim – swum – swum*, with both past tense and past participle identical in <u>. It is not far-fetched to

presume that these three-part verbs are "attracted" in some metaphorical way by the rivalling group of verbs around *cling – clung – clung* or *sling – slung – slung*. This second group of verbs is much rarer in text counts, but many more verb types belong to it than to the group around *drink – drank – drunk*. In other words, the more attractive group wins out by type frequency, although its token frequency is considerably lower.

With the switch of the verb (sub-)class around *drink – drank – drunk* the verbal system is changed significantly. In order to see this, a short digression into the structure of the English verb system is in order here. (For details cf. Anderwald 2009.) English verbs can be classified according to how many distinctions they make between the three "principal forms" INFINITIVE or PRESENT TENSE (abbreviated PRES), PAST TENSE (PAST) and PAST PARTICIPLE (PPL). Five logical possibilities obtain, and all are attested in standard English.

Verb class 1: PRES ≠ PAST ≠ PPL (e.g. *sing – sang – sung*)

Verb class 2: PRES ≠ PAST = PPL (e.g. *fight – fought – fought*)

Verb class 3: PRES = PPL ≠ PAST (e.g. *come – came – come*)

Verb class 4: PRES = PAST ≠ PPL (e.g. *beat – beat – beaten*)

Verb class 5: PRES = PAST = PPL (e.g. *hit – hit – hit*)

However, they are attested to widely varying degrees. Of the "150 or so" (Carstairs-McCarthy 2002:40; in fact in our count more like 167) present-day standard English strong verbs, almost half (49%) are found in verb class 2, which in its abstract pattern is also identical to all weak verbs, a good third (35%) is found in verb class 1, verb classes 3 and 4 are marginal (consisting of all of three verbs, i.e. *come, run* and *beat*), so that the remaining 14% or so of strong verbs pattern with verb class 5, i.e. have no morphological marking for past tense or past participle. Clearly, then, verb classes 1 and 2 deserve some scholarly attention.

Internally, verb class 1 is structured in the following way: nine verbs employ vowel change only (as the cited *sing – sang – sung* or *drink – drank – drunk*), all remaining 50 verbs have a past participle in {en}, in addition to, or instead of, vowel change (cf. *drive – drove – driven, break – broke – broken, fall – fell – fallen, shear – sheared – shorn vs. saw – sawed – sawn*). It is clear that in dialect systems where the first group of verbs around *sing – sang – sung* moves into the neighbouring verb class 2 altogether, this complex system of family resemblances inside verb class 1 is simplified considerably (other things being equal), and the past tense marker {en} is free to become the defining criterion for membership in this verb class (in the sense of Wurzel 1984).

Data from FRED illustrates that past tense levelling to <u> in this groups of new "Bybee" verbs is indeed a sizeable phenomenon, with relative frequencies well beyond the 40 per cent mark for many lexemes, a very high average for a non-standard feature in running text.

	nStE	**StE**	**Sum**	**%**
sink	26	14	40	65.0
drink	18	18	36	50.0
ring	20	25	45	44.4
sing	16	26	42	38.1
begin	11	44	55	20.0
Total	**91**	**127**	**218**	**average 41.7**

Table 2. Non-standard verb forms in FRED

As Bybee has pointed out (1985), all these past tense forms can be described as following the phonological template detailed in (3).

(3) Phonological template $[C\ (C)\ (C)\wedge\{\text{velar/nasal}\}]_{\text{past}}$

This pattern is the most frequent one in the area of English (strong) past tense forms, with 14 verbs following it, and is considerably strengthened in non-standard systems with 23 verbs overall in this group (ca. 14 per cent). It is therefore not surprising that "new" strong verbs (e.g. American English past tense *snuck* or *drug* [for *sneaked* or *dragged*]), other non-standard forms like past tense *run* or *come* (for *ran* or *came*) as well as learners' errors (*brung* for *brought*) also follow this pattern. If we consider the rather disparate nature of English strong verbs overall, this makes for an area of remarkable homogeneity.

Although the template in (3) captures all cases of past tense forms in <u>, the category of Bybee verbs is not without internal structure. We can depict the family resemblances that hold between individual members of this verbal group by way of a two-dimensional grid with the most prototypical members at the centre, and a stepwise change away from a consonant cluster after the vowel on the horizontal axis, and a stepwise change away from initial consonant cluster before the vowel on the vertical axis as in Figure 1. This also makes it graphically clear that, admittedly starting from lexical similarities, these similarities become progressively more dissimilar and more abstract, and indeed more grammar-like. (A rare example of capturing this process in a cognitively plausible way employing stochastic rules is the approach by Albright and Hayes 2003.)

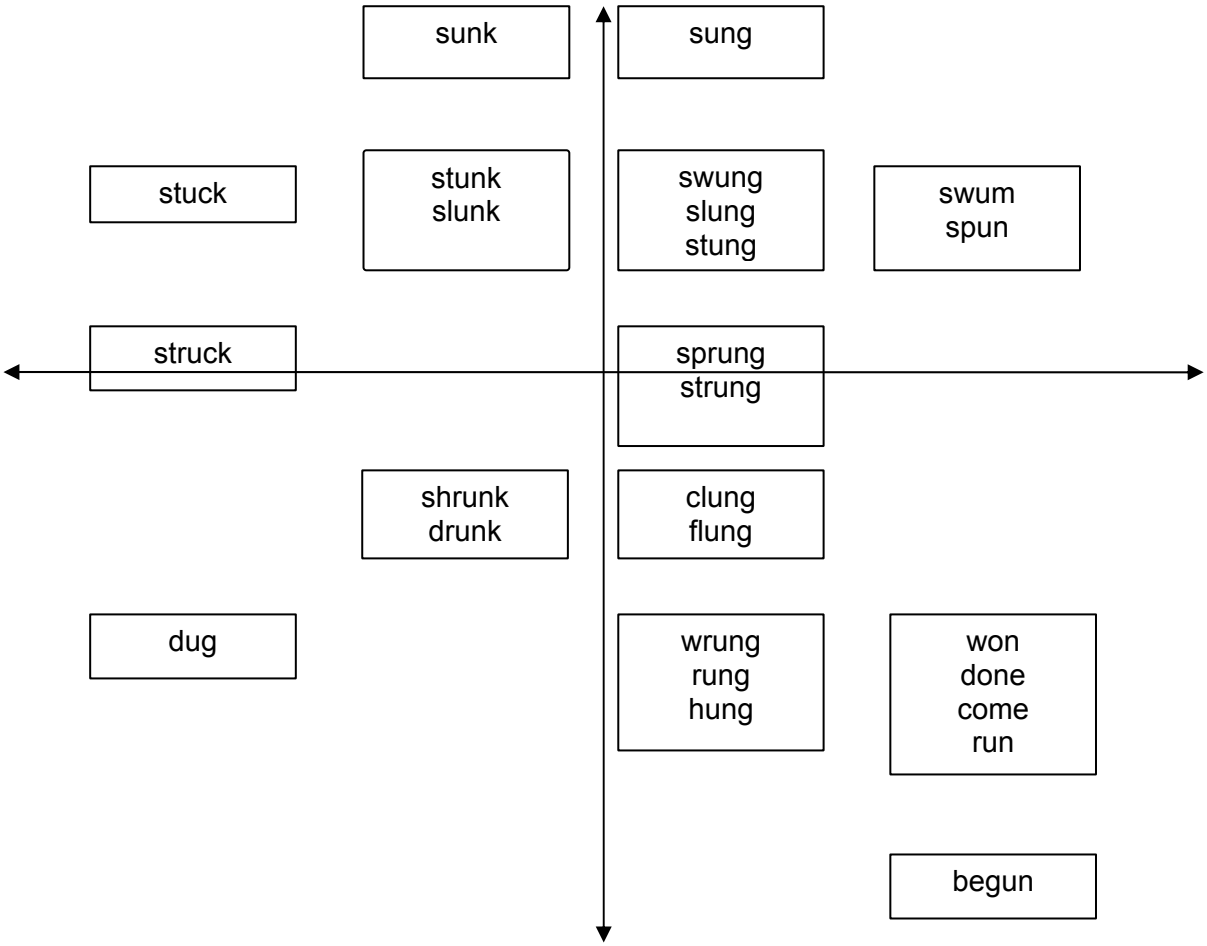

Figure 1. Prototypicality grid of Bybee verbs

Simplifying considerably, the change in dialect systems and the present-day attractiveness of the verb class of Bybee verbs (i.e. verbs around *sling – slang – slung*) indicate that verb classes are real not just for the analyst, but also cognitively real for the language learner. Word-classes that act as an attractor for neighbouring word classes are stable on the basis of extra-morphological characteristics such as the phonological template in (3), and also on the basis of their type frequency. The example of *drink – drank – drunk* has made it clear that although practically all verbs with these three-part paradigms are more frequently encountered in text counts (in terms of token frequency), the group of verbs around *sling – slung – slung* is more numerous in terms of members (i.e. type frequency), and it is clearly type frequency that is the decisive factor when determining the direction of change. This is a conclusion that does not sit easily with purely lexical accounts of grammatical patterns.[1]

4. On the role of markedness and cross-linguistic generalizations

To conclude our overview of grammatical phenomena that can in our view not be explained by the prominent status of lexis alone, this section will discuss one pervasive

[1] For reasons of space, we will not discuss the historical development or Old English word classes, which are clearly relevant in any explanation of this pattern. Anderwald 2009 has devoted a chapter to these issues.

pattern in non-standard negation and relate it to cross-linguistic findings from functional typology.

Apart from well-known features like multiple negation, what we find in all varieties of non-standard English are pervasive patterns of asymmetry. These seem to be particularly pronounced where negation is concerned. Thus, third-person singular *don't* (for *doesn't*), the use of *ain't* for all forms of present tense BE and HAVE or *was/were* levelling to either *wasn't* or *weren't* for all persons (cf. Anderwald 2002 and references therein) have been noted as pervasive dialect features of many, if not most dialects of English. Figures from the British National Corpus for present-day informal spoken English indicate that these features are indeed frequent, as Table 3 shows.

	nStE	StE	sum	% nStE
does + NEG	847	2148	2995	28.3
was + NEG	703	1763	2466	28.5
HAVE + NEG	707	4314	5021	14.1
copula BE + NEG	258	1683	1941	13.1

Table 3. Non-standard negative forms of the primary verbs

The frequent use of non-standard forms under negation (and the relative infrequent, to say the least, use of corresponding positive forms) leads to highly asymmetric paradigms between positive and negative contexts, as figure 2 displays.

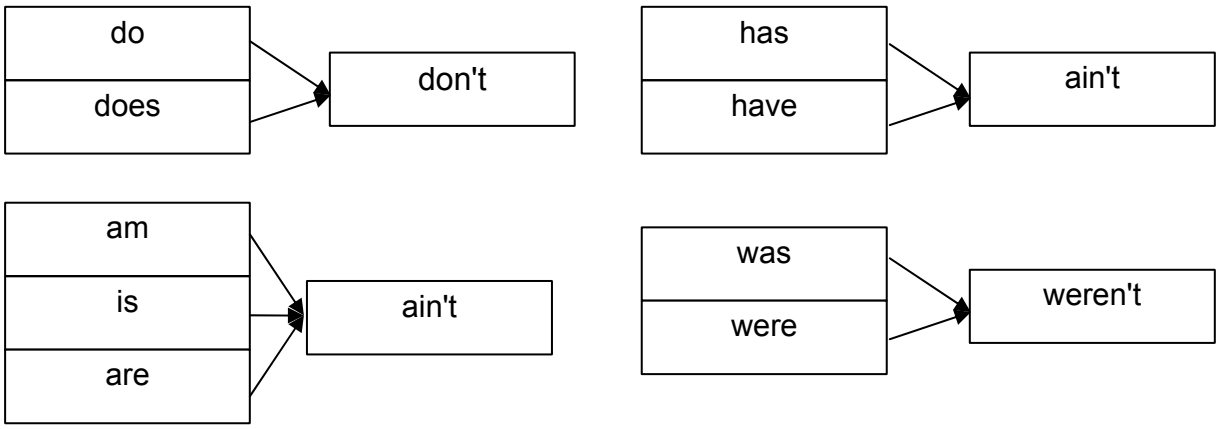

Figure 2. Asymmetric paradigms in non-standard English

This asymmetry between positive and negative context is mirrored by a cross-linguistic preference for more distinctions in positive paradigms than in negative ones. Miestamo has recently interpreted this asymmetry in terms of *vertical analogy*: "the ontology of non-fact is less differentiated than the ontology of fact, and linguistic structure reflects this distinction" (Miestamo 2000: 78). Greenberg and Croft clearly see negation as marked (and positive contexts as unmarked) (Greenberg 1964, Croft

1990); by their criteria, one could see English non-standard negation as 'more marked' vis-à-vis standard English.

	StE	nStE
Zero value	√	√
Syncretization	--	√
Irregularity	--	√
Frequency	√	√

Table 4. Markedness criteria for negation

It is particularly striking that non-standard English paradigms are marked by syncretization, symbolized by merging cells in the negative paradigms above in Figure 2, and by a decrease in irregularity, which is particularly clear in the case of past tense BE. Here, the positive paradigm is the only paradigm across all English verbs that makes a person distinction in the past tense (*was* vs. *were*), but this irregularity is levelled under negation to (regionally different) *wasn't* or *weren't* for all persons.

The fact that we find these asymmetries across very different lexical paradigms in very different phonological environments, mirrored by the same pattern across widely different languages, would speak for more basic cognitive functional reasons that underlie these patterns, rather than see the cause in lexis. Instead, we can observe the re-morphologiziation of a salient cognitive difference, increasing the morphological distance between positive and negative forms, independently of individual lexemes.

5. Conclusion

To summarize, we see grammar as an emergent property diachronically (and perhaps ontogenetically), but certainly not as a synchronic epiphenomenon of lexis. (For a slightly different view, cf. Hopper 1987, 1998.) Instead, we posit a continumm from lexis to patterns to the very abstract patterns of grammar. As we have shown, grammar (as an emergent property) is nevertheless "real", i.e. measurable, and evinces considerable stability, at least synchronically. Although ultimately based in lexis, features of grammar are independent of individual lexemes (sections 2, 4), abstract (sections 2, 3) and may be cognitively motivated (section 4). In this way, we hope to have shown that independently of lexis, grammar still deserves scholarly attention.

References

Albright, Adam, and Bruce P. Hayes (2003): "Rules vs. analogy in English past tenses: A computational/experimental study." *Cognition* 90: 119-161.

Anderwald, Lieselotte (2002): *Negation in Non-Standard British English: Gaps, Regularizations and Asymmetries.* London & New York: Routledge.

Anderwald, Lieselotte (2007): "'He rung the bell' and 'she drunk ale' - non-standard past tense forms in traditional British dialects and on the internet." In Marianne Hundt, Nadja Nesselhauf and Carolin Biewer, eds. *Corpus Linguistics and the Web.* Amsterdam & New York: Rodopi: 271-285.

Anderwald, Lieselotte (2009): *The Morphology of English Dialects: Verb-Formation in Non-Standard English.* Cambridge: Cambridge University Press.

Behaghel, Otto (1909/1910): "Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern". *Indogermanische Forschungen* 25.

Bod, Rens; Hay Jennifer; Jannedy, Stefanie (2003): "Introduction". in: Bod, Rens; Hay Jennifer; Jannedy, Stefanie (eds.): *Probabilistic Linguistics*. Cambridge, MA: MIT Press.

Bybee, Joan. 1985. *Morphology: A Study of the Relation between Meaning and Form.* Amsterdam & Philadelphia: John Benjamins.

Carstairs-McCarthy, Andrew (2002): *An Introduction to English Morphology.* Edinburgh: Edinburgh University Press.

Croft, William (1990): *Typology and Universals.* Cambridge: Cambridge University Press.

Chen, Ping (1986): "Discourse and particle movement in English". *Studies in Language* 10: 79-95.

Greenberg, Joseph (1966): *Language Universals.* The Hague & Paris: Mouton.

Gries, Stefan Th. (2005a): *Multifactorial analysis in corpus linguistics: a study of particle placement*. New York, NY: Continuum.

Gries, Stefan Th. (2005b). "Syntactic Priming: A Corpus-based Approach". *Journal of Psycholinguistic Research* 34: 365-399.

Gries, Stefan Thomas, and Stefanowitsch, Anatol (2004): "Extending collostructional analysis: a corpus-based perspective on 'alternations'". *International Journal of Corpus Linguistics* 9: 97-129.

Hawkins, John (1994): "A performance theory of order and constituency". Cambridge: Cambridge University Press.

Hinrichs, Lars; Szmrecsanyi, Benedikt (2007): "Recent changes in the function and frequency of standard English genitive constructions: a multivariate analysis of tagged corpora". *English Language and Linguistics* 11(3): 437-474.

Hopper, Paul J. (1987): "Emergent Grammar." *Berkeley Linguistics Society* 13: 139-157.

Hopper, Paul J. (1998): "Emergent grammar." In Michael Tomasello, ed. *The New Psychology of Language.* Vol. 1: *Cognitive and Functional Approaches to Language Structure* Mahwah, NJ & London: Lawrence Erlbaum: 155-175.

Kolbe, Daniela; Benedikt Szmrecsanyi (2008): "Complementizer choice in written and spoken English". Paper given at the First Triennial Conference of the International Society for the Linguistics of English (ISLE1), Freiburg, October 2008.

Miestamo, Matti (2000): "Towards a typology of standard negation." *Nordic Journal of Linguistics* 23: 65-88.

Pickering, Martin; Branigan, Holly (1998): "The representation of verbs: Evidence from syntactic priming in language production". *Journal of Memory and Language* 39(4): 633-651.

Stubbs, Michael (2001): *Words and Phrases*. Oxford: Blackwell.

Sankoff, David (1998): "Sociolinguistics and Syntactic Variation" in: Newmeyer, Frederick (ed.): *Linguistics: the Cambridge Survey*, Vol. 4. Cambridge University Press, Cambridge, pp. 140-161.

Szmrecsanyi, Benedikt (2005): "Language users as creatures of habit: a corpus-linguistic analysis of persistence in spoken English". *Corpus Linguistics and Linguistic Theory* 1(1): 113-150.

Szmrecsanyi, Benedikt. (2006): *Morphosyntactic persistence in spoken English: a corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Berlin, New York: Mouton de Gruyter.

Szmrecsanyi, Benedikt; Hernández, Nuria (2007): "Manual of Information to accompany the Freiburg Corpus of English Dialects Sampler ("FRED-S")". Available online at http://www.freidok.uni-freiburg.de/volltexte/2859/. Freiburg: English Dialects Research Group.

Wurzel, Wolfgang U. (1984): *Flexionsmorphologie und Natürlichkeit: Ein Beitrag zur morphologischen Theoriebildung.* Berlin: Akademie-Verlag.