

Against inertia

ABSTRACT

In this paper I question the Inertial Theory of language change put forward by Longobardi (2001), which claims that syntactic change does not arise unless caused and that any such change must originate as an ‘interface phenomenon’. It is shown that these two claims and the contention that ‘syntax, by itself, is diachronically completely inert’ (Longobardi 2001: 278), if construed as a substantive, falsifiable theory of diachrony, make predictions that are too strong, and that they cannot be reduced (as seems desirable) to properties of language acquisition. I also express doubt as to the utility and necessity of a methodological/heuristic principle of Inertia.

Keywords: Diachronic syntax; Inertia; causality; learnability

1. INTRODUCTION

Longobardi (2001) put forward what he termed the INERTIAL THEORY of grammatical change, characterized by the statements in (1–3).

- (1) ‘syntactic change should not arise, unless it can be shown to be *caused* – that is, to be a well-motivated consequence of other types of change (phonological and semantic changes, including the disappearance of whole lexical items) or, recursively, of other syntactic changes’
- (2) ‘linguistic change proper ... may only originate as an interface phenomenon’
- (3) ‘*syntax*, by itself, is diachronically completely inert’
(Longobardi 2001: 277–278; emphasizes his)

This approach draws inspiration from work by Keenan (1994, 2002, 2009), which Longobardi refers to as ‘the pretheoretic concept of *inertia*’ (2001: 277; emphasis his). Keenan expresses this as in (4):

- (4) ‘Things stay as they are unless acted upon by an outside force or DECAY’ (Keenan 2002: 327; emphasis his)

The idea of inertia has received widespread attention in the literature on diachronic generative syntax, as can be seen by the number of papers making reference to it in recent conference volumes (e.g. Ferraresi & Goldbach 2008, Detges & Waltereit 2008, Crisma & Longobardi 2009, Breitbarth, Lucas, Watts & Willis 2010). Reactions have ranged from broadly accepting (e.g. Lightfoot 2002: 130, Hróarsdóttir 2002, 2003,

Ferraresi & Goldbach 2003, Ingham 2006: 257, Roberts 2007: 232, Jäger 2008, Axel & Weiß 2010, Sundquist 2010) to more sceptical (Waltereit & Detges 2008, Biberauer & Roberts 2009: 74, Reintges 2009, Meisel 2011). However, no full critical discussion of the implications of the theory has yet appeared.

In this paper I question the value of the concept of inertia, demonstrating that the notion either is vacuous (trivially/definitionally true) or makes false empirical predictions. Furthermore, I suggest that even as a methodological heuristic the notion of inertia is not necessarily helpful. The version of inertia that I discuss most thoroughly in this paper is the theory spelled out by Longobardi (2001).¹ Keenan's (1994, 2002, 2009) conception of inertia is less susceptible to the criticisms made here, particularly those in section 3, since it is not evidently presented as a hypothesis about the nature of change with empirical consequences.

In section 2 I discuss the intended status of inertia, showing, among other things, that it does not follow from Minimalist assumptions about the nature of the language faculty, and indeed is tangential to the Minimalist Program as ordinarily construed. Section 3 demonstrates that the Inertial Theory as outlined in (1–3), if construed as a substantive, falsifiable hypothesis about language change, *CANNOT* be true given the basic assumptions one must make about the functioning of language acquisition. Section 4 discusses the role of an inertia-based approach as a heuristic, arguing that a methodological principle of this nature is desirable even if trivial; on the other hand, I express some scepticism about its utility and necessity, following Lass (1980, 1997) among others. Section 5 recapitulates and concludes.

¹ It should be noted that I do not wish to question the account of the development of French *chez* that forms the bulk of Longobardi's 2001 paper, which is elegant and supported by rich comparative and philological data.

2. THE INERTIAL THEORY: METHODOLOGICAL, SUBSTANTIVE, MINIMALIST?

It is clear that the Inertial Theory as set out in (1–3), or some combination thereof, cannot be taken as a principle of Universal Grammar or of the faculty of language in a broad sense. This is because the formulation of (1–3) crucially refers to change, which is a relation between different states of the faculty of language in different individuals, and to diachrony. Two options remain: either the Inertial Theory is to be viewed as a principle of change, perhaps reflecting some property of the language faculty, or it constitutes a methodological principle intended for use as a heuristic in investigating change.

Since Longobardi's (2001: 275) stated aim is to 'implement, for diachronic study, the spirit and some guidelines of Chomsky's Minimalist Program', it is worth considering the relation between the Inertial Theory and the Minimalist Program. A common distinction drawn within the Minimalist Program (e.g. by Epstein & Hornstein 1999; Chomsky, Belletti & Rizzi 2002) is that between methodological and substantive Minimalism. Methodological Minimalism 'is simply good scientific practice' (Roberts 2000: 853), and is essentially coextensive with the methodological principle of Occam's Razor: entities must not be multiplied beyond necessity. There is nothing specifically linguistic about such an approach to inquiry. Central to linguistic Minimalism, however, is the question of 'how well designed the system is' (Chomsky, Belletti & Rizzi 2002): substantive Minimalism seeks to address the question of whether the faculty of language itself is optimally designed, as opposed to our theory of it.

One potential assumption is that the Inertial Theory follows from an aspect of substantive Minimalism, namely Chomsky's (1995: 4–5) conjecture that there is no such thing as syntactic variation, with variation 'limited to nonsubstantive parts of the

lexicon and general properties of lexical items'. Under this view, there is no variation across the human species in the syntactic component of the language faculty, and hence no such thing, strictly speaking, as 'syntactic change'.² Instead all change traditionally classed as syntactic is simply lexical change, specifically change in the formal features of (functional) lexical items; this is the approach to variation introduced by Borer (1984) and dubbed the 'Borer-Chomsky Conjecture' (BCC) by Baker (2008: 353). However, statements (1–2) do not support this view of the provenance of the Inertial Theory: if it were hypothesized that there were no such thing as 'syntactic' change, the suggestion that such change does not occur 'unless caused', and that it may originate as an interface phenomenon (preceded by semantic and/or morphophonological change), would be entirely redundant. Moreover, Longobardi (2001) suggests that the theory has 'empirically testable consequences' and that it might turn out to be 'empirically false or only partly correct' (2001: 278). Since Longobardi assumes the BCC (2001: 278), the implication is that the LEXICON is inert, or at least the functional elements within it: if given cases of 'lexical' change have the potential to falsify the Inertial Theory then it cannot simply follow from the BCC that the syntactic component of the language faculty is invariant as discussed above.

Equally, however, if the consequences of the Inertial Theory are empirically testable, then it cannot be intended purely as a heuristic in the sense of methodological Minimalism. The answer to the question posed at the beginning of this section, then, appears to be that (1–3) are to be viewed as a principle of change. However, as

Lightfoot (1979, 1999, 2002) has argued extensively, there can be no principles of

² An exception to this stricture, perhaps, is the question of how the core of the faculty of language itself evolved (cf. the 'evolutionary adequacy' of Longobardi 2003 and Gianollo, Guardiano & Longobardi 2008: 112); but this is a very different question to those usually asked in historical linguistics, and certainly not the one Longobardi (2001) addresses.

history (2002: 134) and ‘there is no theory of change to be had independent of theories of grammar and acquisition’ (2002: 127). It is indeed difficult to see how an ontological claim could be made about diachrony in the same way such claims are made about the faculty of language, since there is no entity of which such a principle could be predicated. It is desirable, then, to reduce the Inertial Theory to properties of the faculty of language and to acquisition; however, I will demonstrate in section 3 that this is impossible given the claims made in (1–3) and the basic assumptions we must make about language acquisition.³

It is also worth raising the question here of why it should be SYNTAX that is inert. Longobardi (2001) takes a neutral position on whether semantics and phonology are inert in this way: he hypothesizes that ‘The semantic and phonological matrices of lexical items will, however, not be similarly constrained’ (2001: 278), adding that ‘It remains to be seen whether the more abstract principles of phonology and morphology are equally subject to the Inertial Theory’ (2001: 278, fn4). However, this seems unlikely, since if we were to hypothesize, following a version of (2), that ALL linguistic change were ‘a well-motivated consequence of other types of change’ (Longobardi 2001: 278) we would predict that there would be no change at all, or at least no change with an ultimately endogenous origin (cf. also the Chicken-and-Egg Problem of Roberts 2007: 125–6). On the other hand, there is no obvious conceptual or empirical reason to assume that syntax is inert but that phonology and morphology are not.

3. WHY AN ONTOLOGICAL INERTIAL THEORY DOES NOT WORK

³ Statement 1, if taken literally, makes little sense, as it implies that the DEMONSTRATION that a syntactic change has a cause, rather than the cause itself, is a necessary condition for that syntactic change to occur. I assume the following reading: ‘syntactic change does not arise unless caused.’

Before an argument against the Inertial Theory can be sketched, some basic assumptions about language acquisition must be laid out. I will start this section by making explicit three such core assumptions in 3.1. The argument against the Inertial Theory itself follows in 3.2 in the form of a thought experiment.

3.1 *Some basic assumptions*

The first assumption I will make is (5).

(5) Assumption of DISCONTINUITY: Acquirers do not have direct access to the grammar of the target language.

(5) as an explicit assumption in work on language change dates back at least as far as Andersen (1973), who schematized language change as in Figure 1:

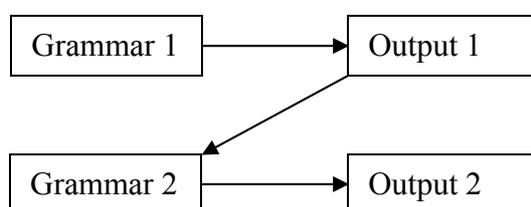


Figure 1

The Z-model of Andersen (1973: 767)

Subsequent work has recognized that matters are rarely quite this simple, and that the primary linguistic data (PLD) that reach the acquirer are usually the product of multiple distinct grammars rather than a single Grammar 1. Longobardi abstracts away from such cases, which he refers to (2001: 278, fn5) as ‘interference’ (cf. also the

discussion of ‘change’ vs. ‘diffusion’ in Hale 1998). The fact that the Inertial Theory can make no predictions in such cases is problematic, since virtually all language acquisition is actually done under conditions of grammar contact of a more or less substantial kind; the Inertial Theory is therefore a theory that holds for an idealized acquisition situation that in fact rarely occurs, and in this respect is similar to the problematic standard assumption made within work on first-language acquisition that the stable state of acquisition corresponds exactly to the target grammar, i.e. that learners converge perfectly on the grammar of their parents/peers (cf. Niyogi & Berwick 1995: 1, Roberts 2007: 229). Even if we were to grant that acquisition based on PLD generated by several virtually identical grammars is different in nature from acquisition based on PLD generated by wildly different grammars (e.g. Modern French and Wolof), and has more in common with acquisition within the idealized homogeneous speech community, ‘interference’ could still occur in these cases with respect to the points in which the similar grammars differ. An example might be the presence vs. absence of ‘British’ *do*, as in *Fred will read the book, and Bill will (do) too*: the presence of *do* here is only grammatical in British English (see e.g. Baltin 2007). Nevertheless, I will follow Longobardi in considering only the change scenario represented by the simplistic model in Figure 1, since this is the type of situation in which syntax is predicted to be inert by (1–3). The essential point is that there is no direct relation between Grammar 1 and Grammar 2; change is instead mediated by Output 1, the PLD, from which the acquirer must infer a grammar. This point is uncontroversial; no one would argue that language learners are telepathic, or that parameter settings are passed down from their parents as part of their genetic

endowment.^{4,5}

The second assumption I wish to highlight is also straightforward if the Principles & Parameters approach to syntactic change is accepted:

- (6) Assumption of EXPERIENCE: There exist parameters for which at least one value requires the presence of positive evidence in the PLD in order to be set.

Assuming (with Longobardi 2001: 278, Borer 1984, Chomsky 1995) that parameters can be reduced to the formal feature specifications of lexical items (see Baker 2008 for discussion), this is an obvious consequence: we might hypothesize that the child requires positive evidence in order to posit the presence of a feature as opposed to its absence.⁶ Longobardi in fact proposes a principle of ‘minimize feature content’ with exactly this effect (2001: 294), representing an aspect of the computational conservatism of the language learner (see Roberts & Roussou 2003 on this notion). More generally, however, (6) is required in order to make parametric acquisition function at all: if neither the presence nor the absence of a given feature requires positive evidence, then both must be entertained simultaneously (against normal assumptions; cf. J. D. Fodor 1998: 21, Roberts 2007: 22), since it is generally assumed

⁴ Though Lightfoot (1979: 391) observes that certain typological theories of long-term drift in fact implicitly assume linguistic ‘racial memories’, and is rightly critical of such approaches.

⁵ I am assuming here that change is intimately linked with first language acquisition, as assumed by Paul (1920), Lightfoot (1979), Hale (1998), Roberts (2007: 123) and, implicitly, Longobardi (2001), among others. The assumption that ALL change stems from L1 acquisition events is far from uncontroversial; however, since this is not the focus of this paper, I abstract away from it here, as does Longobardi (2001).

⁶ As two reviewers observe, this formulation requires that features be privative, whereas much of the syntactic literature assumes a binary [±] or [attribute:value] feature structure. Of course, either of these systems can be reformulated in terms of privative features, though certain feature co-occurrence restrictions may then be required. These complications are purely formal, then, and do not affect the force of the argument here.

that acquirers do not have access to negative evidence (see, for example, Gibson & Wexler 1994: 410 and the references cited there, and Johnson 2004).

For any given parameter, this conclusion can be questioned. One might argue that the posited parameter is not a standalone parameter at all but is rather set as part of a macroparametric cluster (cf. e.g. Baker 2001, 2008) on the basis of different evidence, or that it may be set on the basis of a process such as the ‘generalization of the input’ of Roberts (2007: 275). For instance, the headedness of TP could be hypothesized to be related to a Head Parameter governing the headedness of a much broader set of categories (as proposed by e.g. Chomsky & Lasnik 1993). In both cases, however, positive evidence is still required, in order to trigger the setting of the relevant macroparameter or the input generalization in the first place; (6) must simply hold at one remove from the feature we were initially considering. In the Head Parameter example, it might be argued, for instance, that the headedness of VP is instrumental in setting the headedness of TP.

Even if syntactic change is not framed in terms of parameters but in terms of patterns (Harris & Campbell 1995), rules (Newmeyer 2004, 2005), or constructions, it is still necessary to maintain a version of (6) stating that positive evidence is required in order for syntactic acquisition to take place. In essence, (6) is simply the claim that experience (the second of Chomsky’s (2005) three factors) plays a role in the acquisition of syntax.

The third assumption I wish to highlight is more controversial:

- (7) Assumption of DETERMINISM: The acquisition of syntax is a deterministic process.

Since the term ‘deterministic’ is used in many senses within linguistics, it is useful to clarify this further. The intended meaning of (7) is that, for any temporally ordered set of sentences (PLD), any and all learners exposed to it will converge on the same grammar: there is no “imperfect” learning or “spontaneous” innovation’ (Longobardi 2001: 278).⁷ Clearly (7) is necessary for any version of the Inertial Theory, since imperfect learning and spontaneous innovation cannot be said to be ‘caused’ by interface phenomena: the falsity of (7) entails the falsity of (2). It is therefore also necessary to assume (7) if one wishes to demonstrate that it is logically impossible for the Inertial Theory laid out in (1–3) to be true.

The principle of determinism in (7) is a non-trivial hypothesis about the acquisition of language, and makes empirical predictions that are in principle testable. In fact, the equivalent of (7) for phonology may well be false, since cases of apparently spontaneous innovation with no obvious basis in the PLD have been reported in the literature: one well-known case is the so-called ‘click girl’, a 4-year-old who substituted a dental click for all alveolar and postalveolar affricates and fricatives (Bedore, Leonard & Gandour 1994). Following Hale (1998: 5), we might assume that all such innovations during acquisition should be considered as changes, regardless of whether or not they diffuse. Such innovations could potentially be ascribed to details of the trigger experience, or of perception or the motor organs, of which we are not aware, but the argument for determinism becomes less and less intuitively plausible the further one takes such rescue operations. (7) is also not universally assumed by researchers in syntactic learnability: J. D. Fodor (1998) adopts it, but this is as a reaction to the nondeterministic Trigger Learning Algorithm of Gibson & Wexler (1994), which selects

⁷ The precise definition of PLD varies in the literature, as noted by Hale (1998: 1, fn1). I here take it to be the input to the linguistic learning system in the child’s mind rather than a raw acoustic stream.

a parameter randomly if an unanalyzable sentence is encountered (1994: 410), as do the variants presented by Niyogi & Berwick (1995). The more recent variational model of Yang (2002, 2004: 453) returns to the nondeterministic position by selecting a grammar randomly (with varying probabilities) in order to analyse each input sentence.⁸

Furthermore, if one believes that the language faculty itself matures with age independently of the input received (as do e.g. Borer & Wexler 1992), then we must also stipulate that (7) will only be true if the learners are exposed to sentences in the PLD at the same stage of development. Nevertheless, since I know of no evidence from syntactic acquisition definitively falsifying the assumption of determinism, I take it that (7) is an interesting and valuable hypothesis for researchers in syntactic acquisition and change, since its predictions are strong, and, if true, it would provide an excellent foundation for (comparatively) explanatory accounts of change. Chomsky (1986: 235) in fact implicitly endorses a version of (7) when he states that ‘On the Cartesian assumptions, I attribute to you ... rules that I would have followed had I had your experience.’

It must be noted that (7) is not coextensive with the Inertial Theory, although it is an integral part of it. The Inertial Theory ‘excludes the intervention of probabilistic models in the development of syntax’ (Longobardi 2001: 278), as does (7) insofar as such models include a ‘roll of the dice’ as part of the acquisition algorithm itself, as in the models developed by Clark & Roberts (1993), Gibson & Wexler (1994) and Yang (2002, 2004). However, unlike (1–3), (7) by itself makes no claim about the relation between grammars diachronically: it does not claim that acquirers will converge

⁸ Yang (2002: ch. 5) in fact takes a similar position to (7) in his discussion of diachrony: ‘the only source for the discrepancy between two generations of speakers must lie in the linguistic evidence’ (2002: 127). However, this does not follow as a consequence of his model, which is probabilistic, as discussed.

perfectly on the target grammar, and in this respect is more similar to the ‘weak determinism’ of Roberts & Roussou (2003: 13) and Roberts (2007: 231). Given that the Inertial Theory in (1–3) is false if the assumptions of discontinuity and experience in (5–6) are true, as demonstrated in section 3.2, I tentatively suggest that the principle of determinism in (7) is methodologically the ‘next best thing’ to the Inertial Theory in terms of empirical predictions about diachrony.

3.2 *The thought experiment*

Imagine a child whose parents’ grammar, Grammar 1, requires V-to-C movement in *wh*-questions. I take it here that the postulation of the presence of this movement requires evidence in the data, following the assumption of experience, (6); this may not in fact be the case for this feature, of course. The crucial point is that at least one aspect of the parents’ grammar is not acquirable without positive evidence, and here, for the purpose of exemplification, I take one such aspect to be V-to-C movement in *wh*-questions.

Like Longobardi (2001), I abstract away from ‘interference’ and assume that the PLD the child is exposed to is generated by a single grammar, or at least uniform with respect to this particular feature. Slightly implausibly, but for the sake of concreteness, let us suppose that the child is only exposed to sentences spoken by her parents, both of whom have identical grammars in all relevant respects.

Now let us suppose that the parents never needed or wanted to ask direct questions in the presence of the child, and therefore that Output 1—the PLD—includes no relevant examples. The child therefore fails to acquire V-to-C movement in *wh*-questions in her grammar, Grammar 2.⁹

⁹ Of course, another logical possibility is that the property requiring positive evidence is V-in-situ rather than V-to-C movement. In this case, the argument could be constructed precisely the opposite way round, with Grammar 1 lacking V-to-C movement and

Syntactic change has clearly occurred in the above scenario. Note that there is no change in Grammar 1 itself, only in its output, which cannot be attributed to intralinguistic factors. Yet Grammar 2 is different. Is this change ‘caused’, in the terminology of (1)? On the standard scientific assumption that the world is causally structured (cf. Popper 1968: 67 and section 4.1 below), it is desirable to assume that it is. Yet this cause cannot be said to be a well-motivated consequence of other types of change; instead the cause is clearly an extralinguistic one, namely whatever motivated the fluctuation in the trigger experience. Here it is essentially chance that has ‘caused’ the change; even assuming (7), there is simply no guarantee that the PLD will contain relevant examples.

The non-expression of a particular type of datum in the PLD is, of course, itself causal in a sense: it is what Lightfoot (2006: 165) terms the ‘local cause’ of a change. However, as (1) makes clear, this is not what Longobardi (2001) conceives of as a ‘cause’. Furthermore, the acceptance of non-expression in the PLD as ultimate cause of a change is not conducive to a satisfactorily explanatory model of syntactic change, since the details of the precise PLD available to acquirers at a specific point in time are not within our reach (as Lightfoot (2006: 159) observes), and so any such ‘explanation’ will be *post hoc* and stipulative. The natural next question to ask is what motivated this non-expression, and in order to answer this question extralinguistic notions must be taken into consideration.

To take the scenario in a slightly different direction, it could have been the case that the parents were members of a religious group whose teachings state that it is deeply sinful to ask direct questions. Being good cult members, the parents adhered to Grammar 2 innovating it. Moreover, a reviewer notes the intriguing possibility that in this scenario the child would fail to acquire *wh*-questions entirely; this would, of course, involve a change in any case.

this restriction in an exemplary fashion, and so in this case too there would be no relevant examples in the PLD. If we are to maintain (1), the notion of causality must be understood to be so broad as to be entirely vacuous, making no empirical predictions at all, since a wide variety of extralinguistic events, and even chance non-occurrence, would need to be analysable as a ‘cause’. At the very least, human intentionality—the ‘planning’ of utterances—must be taken into account, and it has been argued that this is inherently non-deterministic (Popper & Eccles 1977; Lass 1980: 31) and therefore not amenable to explanation in terms of causality.¹⁰ Lightfoot (2006: 165) makes a similar point: ‘What we cannot explain, in general, is why the linguistic environment should have changed in the first place’, since a vast number of contingent factors are involved. All in all, then, it seems that (1), the claim that syntactic change does not arise unless caused, either is axiomatic and trivially true, hence making no empirical predictions, or makes predictions that are too strong.

The scenario above is even more of a problem for (2), since there can be no question that the change in this scenario might have originated as an ‘interface phenomenon’: no semantic or morphophonological change preceded it. The change is purely syntactic, involving only the formal features of items in the lexicon. The only way to deny this would be to claim that the syntactic parameter in question (V-to-C movement in *wh*-questions, or any other example one cares to imagine, hence *all* syntactic parameters) were set entirely on the basis of semantic and/or

¹⁰ Which is not to say that all aspects of human cognition are so; for instance, it seems plausible that the acquisition process, and the process of syntactic derivation, might be ‘mechanical’ in the requisite sense, with no intentionality directly involved, and indeed this hypothesis is fundamental to cognitive science as a discipline. But extending this type of determinism to intentionality means that ‘we must end up as behaviourists, and reduce man to a *bête-machine*’ (Lass 1980: 102). Such a stance would not be compatible with the mentalist view adopted by e.g. J. A. Fodor (1983) and Chomsky (1975: ch4, 1995: 2), among others, that the study of the ‘central system’ of the mind, as opposed to its submodules, is an intractable problem.

morphophonological features, and that the absence of these features in the PLD was responsible for their non-acquisition in the child's grammar and thus indirectly for the syntactic change. But as well as reducing the whole Inertial Theory to vacuous triviality, this equates to denying (6), since by claiming that all syntactic (formal) features are set on the basis of semantic and/or morphophonological features one is denying the existence of syntactic acquisition as a primitive process. As discussed in 3.1, such a position is untenable, and not the one adopted in Longobardi (2001). Finally, if (2) is false of this scenario, then (3) is also false: syntax is not 'diachronically completely inert', since we have here a syntactic change which occurred apparently without a semantic or morphophonological cause.

Some further comment is needed on the nature of the scenario mentioned above. It involves *SYSTEMATIC* rather than random endogenous variation, in that the PLD is not purely a random subset of tokens of grammatical sentences as assumed in work on learnability (e.g. by Niyogi & Berwick 1995: 10). The assumption of randomness may be a necessary step in a formal model of acquisition in order to avoid incorporating a Theory of Everything into the model, but there is no reason to believe that it reflects any underlying reality. Parents are perfectly capable of choosing to lock their child in the cellar and to address it using sentences composed solely of the word *fish*, and such engineered systematic endogenous variation would be bound to have implications for the grammar acquired by the child. Assuming random endogenous variation, however, one cannot guarantee that all sentences will *NOT* be composed solely of the word *fish*. Hence the (realistic) assumption of the possibility of systematic endogenous variation is not even necessary for the scenario above to go through.

One may object that the scenario sketched in this section is incredibly unlikely

to occur. This should not affect the force of the argument, however, since the logical possibility of such a change cannot be denied. As Niyogi & Berwick (1995: 2) put it, ‘even if the PLD comes from a single target grammar, the actual data presented to the learner is truncated, or finite. After a finite sample sequence, children may, with non-zero probability, hypothesize a grammar different from that of their parents.’ In other words, models of language learning that meet Gold’s (1967) learnability criterion of ‘identification in the limit’ cannot be taken to lead to successful convergence in real-world cases of language acquisition, as Gold himself recognizes (1967: 450; see also Johnson 2004). Even if it is the case that the overwhelming majority of syntactic changes originate as interface phenomena (and can be said to be ‘caused’ in an explanatorily useful sense), the scenario outlines a counterexample to (2–3), which as a consequence can no longer be upheld as universal claims about diachrony, at least not if they have any predictive power. We are therefore forced to conclude that the Inertial Theory, in its substantive form, is false.

4. INERTIA AS A METHODOLOGICAL PRINCIPLE

As well as claiming that the Inertial Theory has ‘empirically testable consequences’, Longobardi (2001: 278) refers to it as a ‘research program’, adding that ‘even if the Inertial Theory turns out to be empirically false or only partly correct, an important quality ... is its heuristic value: it forces us to look for explanations for all syntactic changes’. This section is devoted to exploring the consequences of this methodological principle. Explanations for syntactic (and other) changes are, obviously, desirable: such a position is so common as to be trivial among historical linguists. However, this section argues that it is not necessarily the case that all our energies should be devoted to

seeking such explanations, and that it is certainly not the case that an approach that fails to do so is unscientific or uninteresting. Relevant considerations to this end are drawn from three different disciplines: historical phonology in 4.1, the study of syntactic learnability in 4.2, and (more tentatively) evolutionary biology in 4.3.

It should be noted here that it is easy to see why such a methodological principle might be proposed. As Longobardi notes (2001: 275, emphasis his), ‘much recent work in diachronic syntax has actually been guided by the aim of *describing* changes (e.g., parameter resetting), rather than by concerns of genuine *explanation*.’ Here I concur: although formalizing a syntactic change in terms of parametric change is a valuable exercise in itself, it is not diachronically explanatory alone. However, a general theory of how internally caused change (as illustrated in section 3.2) can arise is at least as valuable as a set of external explanations of particular changes, and for this reason if no other we should be wary of a methodological Inertia principle. Furthermore, once the possibility of internally caused change in general is accepted, then methodological Minimalism gives us no reason to suspect that any particular syntactic change should be ‘caused’ by preceding semantic or morphophonological changes in the sense of the Inertial Theory.

4.1 *Explanations in historical phonology*

The stance taken by Longobardi (2001) contrasts markedly with that taken in historical phonology, a field in which, arguably, few if any true explanations of individual changes have ever been proposed. Although the Neogrammarians, to whose work Longobardi compares his own (2001: 278), identified broad areas which could be responsible for language change in general, such as child language acquisition (Paul 1920 [1880]) and

ease of articulation, they came up with no specific causal explanations that are accepted today; cf. McMahon (1994: 18) for discussion.

Much subsequent work has made a great deal of progress in identifying the sort of circumstances in which (certain types of) changes typically occur. To take just a few examples, the work of Ohala (e.g. 1981) has shown that abrupt phonological changes can typically occur when the speech signal is reanalysed by the listener, e.g. the nasalization of vowels before nasal consonants and subsequent loss of the consonant itself. Blevins (2004: 32) sets out three possible routes by which sound change can take place: misperception, misapplication of mapping from phonetic to phonological form in ambiguous inputs, and choice of a different variant as prototype from that of Grammar 1 (cf. Bowerman 2008: 198 for an interesting attempt to extend this model to syntax). Labov (2007) has proposed a distinction between change located in first language acquisition (transmission) and change located in ‘extragenerational learning’ (diffusion), with the latter typically being ‘slower, less regular, and less governed by structural constraints’ (2007: 383), exemplified by the incomplete outward spread of the complicated short-*a* tensing system of New York City (2007: 353–372).

Yet there is a sense in which all the above works can be taken as proposing typologies of change situations rather than actually explaining changes. This is the position taken by Lass (1980), who, after examining and dismantling a wide range of candidates for explanations in the causal, deductive-nomological sense, argues that it may not be appropriate to seek such explanations in linguistics. The central conclusion of Lass (1980) can be summed up by the claim that ‘there are no D[educative]-N[omological] explanations in historical linguistics, because “laws” of the appropriate kind do not exist’ (1980: 90). Lass also argues (1980: 101–3) that Popper’s

methodological version of the principle of causality, ‘the simple rule that we are not to abandon the search for universal laws and for a coherent theoretical system, nor ever give up our attempts to explain causally any kind of event we can describe’ (Popper 1968: 67), while attractive, may not be appropriate ‘for any discipline whose main interest is in the behaviour of sentient beings’ (1980: 102). Lass (1997), while acknowledging some weaknesses of his earlier work, remains convinced that it is not the case that ‘causal explanations are or ever will be available’ in historical linguistics (1997: 336). At the core of his contention is the argument that human intentionality must be involved in language transmission/diffusion (a view also endorsed in section 3.2 of this paper), that such intentionality should not be described in terms of nomic causality, and that such causality is therefore not an appropriate concept in studying language change. Discussion of this important thesis is notably absent from Longobardi’s 2001 paper. Since the heuristic version of the Inertial Theory mentioned at the beginning of this section bears strong similarities to Popper’s methodological version of the principle of causality as outlined above, which Lass claims is inappropriate to diachronic linguistics, this is a worrying omission.

In conjunction with the criticisms presented in section 3 of this paper, there is indeed reason to doubt that the Inertial Theory should be accepted even as a heuristic. While it may be the case that, given a set of PLD and an appropriate learning algorithm, we can predict the mature state of an individual’s linguistic competence (cf. (7) above) or perhaps the probabilities of different linguistic states being reached, we have seen that it is impossible for us to predict this state based solely on the grammar(s) generating the PLD, since the PLD itself is just too unpredictable. Given this result, there is no reason to expect that, following a heuristic version of the Inertial Theory, one

would find the sort of linguistic ‘cause’ discussed in Longobardi (2001), even for *post hoc* explanations of changes that have occurred. Here, therefore, one is broadly led to concur with Lass (1980: 99–101) that such a heuristic, while attractive, is inappropriate in historical linguistics, or at least that it should not be the only direction taken.

4.2 *Explanations in syntactic learnability theory*

Inertia as a heuristic is also strangely at odds with the results of syntactic learnability theory, as hinted at in section 3.1. Briefly put, many learnability algorithms proposed in the literature make no claims about relations between Grammar 1 and Grammar 2 in the way that the ontological version of the Inertial Theory does, and nor do they even assume (7)—that language acquisition is a deterministic process. The Trigger Learning Algorithm of Gibson & Wexler (1994), upon which the diachronic model of Niyogi & Berwick (1995) is based, relies on the existence of local maxima to explain change: the learning algorithm contains a ‘roll of the dice’ which may lead learners irretrievably astray in a certain proportion of cases. Similarly, the probabilistic component of the model developed by Yang (2002) may lead to the acquirer assigning different weights to certain hypotheses than the individuals from whose competence the PLD is generated. In the account of language learning as a genetic algorithm in Clark & Roberts (1993), ‘since nothing in the approach requires ... [the fittest] grammar to be consistent with the one that underlies the input text, learners may arrive at final-state systems that differ from those of their parents’ (1993: 303). Their algorithm contains a ‘mutation operator’ which alters the value of a randomly selected parameter (1993: 310–311).

Since it is agreed among researchers in diachronic generative syntax (following Lightfoot 1979) that properties of change should ideally fall out from a theory of

grammar and a learning algorithm, there is a strong case to be made that the assumptions underlying these models should not be rejected out of hand. And, just as we found in the last section, these assumptions conflict with those of the Inertial Theory: in none of these models is it possible to discover a necessary and sufficient condition for EVERY change that might occur.¹¹ Even in a model such as that of J. D. Fodor (1998), which does accept a deterministic process of acquisition, there can be no guarantee that the actual PLD will provide the necessary syntactic evidence for convergence in parameter setting. It follows that looking for such causes should not be the be-all and end-all of diachronic syntax as methodological Inertia would have it.

4.3 *Explanations in evolutionary biology*

This last subsection is devoted to exploring the parallels between a domain of diachronic biology and historical linguistics (not a new endeavour; cf. Paul 1920 [1880], Lightfoot 1979, Lass 1980: 103–109, Clark & Roberts 1993 and Yang 2002, to name but a few).

As pointed out by Clark & Roberts (1993: 300–301), the sequence of parameter settings representing a ‘grammar’ (and thus, broadly speaking, an individual’s competence in a language, abstracting away from competing grammars in the sense of Kroch 1989 and Yang 2002) can be taken to be analogous to an individual’s DNA. Importantly, like grammars, DNA is not transmitted directly from individual to individual or from cell to cell: in the case of replication, its transmission is mediated by

¹¹ As an aside, a reviewer notes that acceptance of the BCC and of a microparametric approach to syntactic change raises a potentially serious problem with regard to learnability, since learning algorithms of the kind discussed in this section have their formal basis in traditional parametric theory, and it is not immediately obvious how they can be expressed in microparametric terms. This is clearly an area in which substantial research is needed.

DNA polymerase enzymes, a process that can be taken to be analogous to language acquisition in this context. As in the case of language acquisition, the fidelity of this process is very high: the new strand matches its template in the vast majority of cases, in the same way that the acquirer's grammar matches that which generated the PLD. However, mistakes do occur, roughly once for each 10^9 nucleotides copied (Alberts *et al.* 2002: ch. 5): mutations arising from copying error in this way are generally described as spontaneous. Spontaneous mutations could be regarded, in turn, as analogous to the type of linguistic change whose possibility was outlined in section 3.2, in which variation in the data actually produced from Grammar 1, for extralinguistic reasons, makes it impossible for the acquirer to match it in Grammar 2. It is important to note that neither in the case of 'spontaneous' linguistic change nor in the case of 'spontaneous' mutation are we forced to conclude that the event has NO cause at all that can be incorporated into a deterministic model (though one might conclude this; cf. Mayr 1968 and the discussion in Lass (1980: 106–107)). Rather, we might make a methodological decision to ignore the causes of such events because they take place on a level in which we are not interested or to which we have insufficient access: in the biological case, molecular decay;¹² in the linguistic case, the precise distribution of the PLD accessible to a single acquirer. These levels might be amenable to investigation if we were to conduct the right (synchronic) study; however, for specific historical developments we have no access to this information either in linguistics or in biology.

¹² This recalls Keenan's original formulation of Inertia given in (4) above, which makes reference to decay. Indeed, a notion of decay seems to be what is missing from Longobardi's (2001) formulation. Accepting (7), that language acquisition is (weakly) deterministic, one may wish to view decay as driven by the ('accidental') non-occurrence of relevant data in the PLD, which leaves us with a form of 'weak inertia' diachronically, as suggested to me by Ian Roberts (p.c.). This 'weak inertia', if one wants to call it that, is coextensive with the assumption of determinism in (7) and compatible with Keenan's (2002) view of Inertia, but not with the theory presented in Longobardi (2001).

This has been the approach taken in evolutionary biology since Darwin: as Lewontin (1983: 65–66) stresses, the emphasis in this variational paradigm has been on explaining why certain types persist and others do not, with the question of how variants themselves arise accorded only secondary importance if considered at all. I would suggest that the task of the historical syntactician should be the same as that of the Darwinian evolutionary biologist: to explain, through reference to endogenous factors (such as the Transparency Principle of Lightfoot 1979, or the Subset Principle of Berwick 1985, Atkinson 2001, Biberauer & Roberts 2009, to name just a few potential examples) or exogenous factors (language contact etc.) as appropriate, the persistence and spread of types once those types have come into being. This conclusion runs counter to what has been suggested by Hale (1998), among others, and yet I believe it to be the only sensible one, given that the specifics of the PLD of acquirers of previous millennia are forever beyond our grasp. Hale's assertion that 'diffusion ... represents the trivial case of acquisition: accurate transmission' (1998: 5) is highly problematic. In the vast majority of cases of diffusion, the acquirer will be faced with data from several distinct grammars, not just one, and must choose between them (again abstracting away from competing grammars in the sense of Kroch 1989 or Yang 2002). Exactly how this choice is made is a nontrivial question, and I see no reason that endogenous as well as exogenous factors should not be involved. Willis (1998: 47–48) in fact proposes such a model, in terms of 'multiple reactivation', where the same actuation triggered by the same factors occurs in multiple speakers, rather than diffusion alone. Importantly, I do not mean to suggest that we should abandon the close link between acquisition and language change emphasized by Lightfoot (1979, 1999, 2006). In fact, if anything, such a 'multiple reactivation'-based approach brings more actual instances of language

change into the purview of an acquisitionist approach to change, since recourse to endogenous factors may be taken to explain cases that Hale (1998) and other narrow actuationists would class as ‘diffusion’ and as belonging to a different explanatory domain.

It is not the case that one approach is *a priori* more valuable or scientific than the other. The actuation-centred approach advocated by Longobardi (2001) and Hale (1998, 2007) is perhaps more in the spirit of (an idealized version of) physics, while the approach I am advocating bears more affinities with evolutionary biology (cf. also Lightfoot 2006: 165). To an extent, the two approaches are complementary; however, I concur with Lass (1980) in arguing that the latter approach may be more methodologically appropriate for historical linguistics, given its subject matter, and certainly for the explanation of specific changes. Rather than seeking narrowly causal explanations of a kind which may be impossible to achieve, we should instead be identifying the factors (both endogenous and exogenous) which might have aided a variant grammar in persisting or becoming more prevalent.

5. CONCLUSION

I began this paper by introducing inertia and the Inertial Theory and clarifying some of the basic notions behind it, including its relation to the Minimalist Program (sections 1–2). I then demonstrated that, under reasonable assumptions about syntactic acquisition, the ontological Inertial Theory makes predictions that are too strong (section 3). In the process I suggested an alternative that may be worthy of further consideration: the slightly weaker notion that the acquisition of syntax is a deterministic process. Finally, and more speculatively, I offered some suggestions as to why the methodological

heuristic correlate of the Inertial Theory may not be the ideal guiding force in our field, drawing on neighbouring disciplines for illustration (section 4), and outlined my view of an ideal diachronic syntax.

It may be objected at this stage that we have come a long way from the original intuition behind Inertia: the simple and appealing notion, most elegantly captured by Keenan (1994, 2002, 2009), that linguistic change is not wildly unconstrained.

However, such an objection is orthogonal to the main concern of this paper.

Longobardi's (2001) specific formulation of an Inertial Theory, he claims, is a 'nontrivial hypothesis' and has 'empirically testable consequences' (2001: 278); this is very different from the inoffensive intuition just discussed. The main aim of this paper has been to show that, if the Inertial Theory is really a nontrivial hypothesis, it is a false one; a subsidiary aim of this paper has been to show that as a 'research program' and as a 'heuristic' (2001: 278) it may not be the only, or indeed the best, way to make progress in the field of diachronic syntax. The original intuition behind inertia thus stands unsullied.

We should, then, seek alternatives to the Inertial Theory. In its place, I would suggest, we have two research directions. One is the conjecture that the acquisition of syntax, or perhaps language acquisition more generally, is a deterministic process. The other is the notion that the diffusion of linguistic variants across populations should be given a much more central position in current theorizing in diachronic syntax, in an approach that takes first language acquisition and endogenous factors to play a key role in diffusion/reactuation as well as in traditional actuation, broadly following the research tradition initiated by Lightfoot (1979). Both, I hope, represent interesting avenues for exploration.

ACKNOWLEDGEMENTS

Parts of this work were presented at SyntaxLab in Cambridge and at the Sixth Cambridge Postgraduate Conference in Linguistics (CamLing), 2010. I am grateful to audiences there and to other readers, particularly Tim Bazalgette, Theresa Biberauer, Chris Lucas, Ian Roberts, David Willis and three anonymous reviewers for *Lingua*, for their comments and suggestions. It goes without saying that none of these people necessarily agrees with my stance, and that any errors are mine alone.

REFERENCES

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. 2002. *Molecular biology of the cell*. Garland Science, New York.
- Andersen, H., 1973. Abductive and deductive change. *Language* 49, 765–793.
- Atkinson, M., 2001. Learnability and the acquisition of syntax. In: Bertolo, S. (Ed.), *Language acquisition and learnability*. Cambridge University Press, Cambridge, pp. 15–80.
- Axel, K., Weiß, H., 2010. What changed where? A plea for the re-evaluation of dialectal evidence. In: Breitbarth et al. (Eds.), pp. 13–34.
- Baker, M., 2001. *Atoms of Language: The Mind's Hidden Rules of Grammar*. Basic Books, New York.
- Baker, M., 2008. The macroparameter in a microparametric world. In: Biberauer, T. (Ed.), *The limits of syntactic variation*. John Benjamins, Amsterdam, pp. 351–373.
- Baltin, M., 2007. The non-unity of VP-preposing. *Language* 82, 734–766.
- Bedore, L., Leonard, L., Gandour, J., 1994. The substitution of a click for sibilants: a case study. *Clinical Linguistics and Phonetics* 8, 283–293.
- Biberauer, T., Roberts, I., 2009. The return of the Subset Principle. In: Crisma & Longobardi (Eds.), pp. 58–75.
- Blevins, J., 2004. *Evolutionary phonology*. Cambridge University Press, Cambridge.
- Bloomfield, L., 1927. Review of J. O. H. Jespersen, *The Philosophy of Grammar*. *Journal of English and Germanic Philology* 26, 244–246.
- Borer, H., 1984. *Parametric syntax*. Foris, Dordrecht.
- Borer, H., Wexler, K., 1992. Bi-unique relations and the maturation of grammatical

- principles. *Natural Language and Linguistic Theory* 10, 147–189.
- Bowern, C., 2008. Syntactic change and syntactic borrowing in generative grammar. In: Ferraresi & Goldbach (Eds.), pp. 187–216.
- Breitbarth, A., Lucas, C., Watts, S., Willis, D. (Eds.), 2010. *Continuity and change in grammar*. John Benjamins, Amsterdam.
- Chomsky, N., 1975. *Reflections on Language*. Pantheon, New York.
- Chomsky, N., 1986. *Knowledge of Language: its nature, origin, and use*. Praeger, New York.
- Chomsky, N., 1995. *The Minimalist Program*. MIT Press, Cambridge, MA.
- Chomsky, N., 2005. Three factors in language design. *Linguistic Inquiry* 36, 1–22.
- Chomsky, N., Belletti, A., Rizzi, L., 2002. *On nature and language*. Cambridge University Press, Cambridge.
- Chomsky, N., Lasnik, H., 1993. Principles and Parameters theory. In: Jacobs, J., von Stechow, A., Sternefeld, W., Vennemann, T. (Eds.), *Syntax: an international handbook of contemporary research*. Walter de Gruyter, Berlin/New York, pp. 506–569.
- Clark, R., Roberts, I., 1993. A computational model of language learnability and language change. *Linguistic Inquiry* 24, 299–345.
- Crisma, P., Longobardi, G. (Eds.), 2009. *Historical syntax and linguistic theory*. Oxford University Press, Oxford.
- Detges, U., Waltkeit, R. (Eds.), 2008. *The paradox of grammatical change: perspectives from Romance*. John Benjamins, Amsterdam.
- Epstein, S. D., Hornstein, N., 1999. Introduction. In: Epstein, S. D., Hornstein, N. (Eds.), *Working minimalism*. MIT Press, Cambridge, MA, pp. ix–xviii.

- Ferraresi, G., Goldbach, M., 2003. Some reflections on inertia: Infinitive complements in Latin. In: Baumgarten, N., Böttger, C., Motz, M., Probst, J. (Eds.), *Übersetzen, Interkulturelle Kommunikation, Spracherwerb und Sprachvermittlung - das Leben mit mehreren Sprachen: Festschrift für Juliane House zum 60. Geburtstag*, pp. 1–12. [Zeitschrift für Interkulturellen Fremdsprachenunterricht [Online] 8.]
- Ferraresi, G., Goldbach, M. (Eds.), 2008. *Principles of syntactic reconstruction*. John Benjamins, Amsterdam.
- Fodor, J. D., 1998. Unambiguous triggers. *Linguistic Inquiry* 29, 1–36.
- Fodor, J. A., 1983. *The modularity of the mind*. MIT Press, Cambridge, MA.
- Gianollo, C., Guardiano, C., Longobardi, G., 2008. Three fundamental issues in parametric linguistics. In: Biberauer, T. (Ed.), *The limits of syntactic variation*. John Benjamins, Amsterdam, pp. 109–142.
- Gibson, E., Wexler, K., 1994. Triggers. *Linguistic Inquiry* 25, 407–454.
- Gold, E. M., 1967. Language identification in the limit. *Information and Control* 10, 447–474.
- Hale, M., 1998. Diachronic syntax. *Syntax* 1, 1–18.
- Hale, M., 2007. *Historical linguistics: theory and method*. Blackwell, Oxford.
- Harris, A., Campbell, L., 1995. *Historical syntax in cross-linguistic perspective*. Cambridge University Press, Cambridge.
- Hróarsdóttir, T., 2002. Explaining language change: a three-step process. *Linguistics in Potsdam* 19, 103–141.
- Hróarsdóttir, T., 2003. Language change and language acquisition. *Nordlyd* 31, 133–155.

- Ingham, R., 2006. On two negative concord dialects in early English. *Language Variation and Change* 18, 241–266.
- Jäger, A., 2008. *History of German negation*. John Benjamins, Amsterdam.
- Johnson, K., 2004. Gold's theorem and cognitive science. *Philosophy of Science* 71, 571–592.
- Kayne, R., 1994. *The antisymmetry of syntax*. MIT Press, Cambridge, MA.
- Keenan, E., 1994. *Creating anaphors: an historical study of the English reflexive pronouns*. Ms., University of California at Los Angeles.
- Keenan, E., 2002. Explaining the creation of reflexive pronouns in English. In: Minkova, D., Stockwell, R. (Eds.), *Studies in the history of the English language: a millennial perspective*. Mouton de Gruyter, Berlin/New York, pp. 325–354.
- Keenan, E., 2009. Linguistic theory and the historical creation of English reflexives. In: Crisma & Longobardi (Eds.), pp. 17–40.
- Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1, 199–244.
- Labov, W., 2007. Transmission and diffusion. *Language* 83, 344–387.
- Lass, R., 1980. *On explaining language change*. Cambridge University Press, Cambridge.
- Lass, R., 1997. *Historical linguistics and language change*. Cambridge University Press, Cambridge.
- Lewontin, R., 1983. The organism as the subject and object of evolution. *Scientia* 118, 65–82.
- Lightfoot, D., 1979. *Principles of diachronic syntax*. Cambridge University Press,

- Cambridge.
- Lightfoot, D., 1999. *The development of language: acquisition, change and evolution*. Blackwell, Oxford.
- Lightfoot, D., 2002. Myths and the prehistory of grammars. *Journal of Linguistics* 38, 113–116.
- Lightfoot, D., 2006. *How new languages emerge*. Cambridge University Press, Cambridge.
- Longobardi, G., 2001. Formal syntax, diachronic Minimalism, and etymology: the history of French *chez*. *Linguistic Inquiry* 32, 275–302.
- Longobardi, G., 2003. Methods in parametric linguistics and cognitive history. *Linguistic Variation Yearbook* 3, 101–138.
- Mayr, E., 1968. Cause and effect in biology. In: Waddington, C. (Ed.), *Towards a theoretical biology, 1: prolegomena*. An IUBS symposium. Edinburgh University Press, Edinburgh, pp. 42–54.
- McMahon, A., 1994. *Understanding language change*. Cambridge University Press, Cambridge.
- Meisel, J. M., 2011. Bilingual language acquisition and theories of diachronic change: bilingualism as cause and effect of grammatical change. *Bilingualism: Language and Cognition* 14, 121–145.
- Newmeyer, F., 2004. Against a parameter-setting approach to typological variation. *Linguistic Variation Yearbook* 4, 181–234.
- Newmeyer, F., 2005. *Possible and Probable Languages: A Generative Perspective on Linguistic Typology*. Oxford University Press, Oxford.
- Niyogi, P., Berwick, R., 1995. The logical problem of language change. MIT Artificial

- Intelligence Laboratory Memo No. 1516. Cambridge, MA.
- Ohala, J., 1981. The listener as a source of sound change. In Masek, C., Hendrick, R., Miller, M. (Eds.), *Papers from the Parasession on Language and Behavior*. Chicago Linguistic Society, Chicago, pp. 178–203.
- Paul, H., 1920 [1880]. *Prinzipien der Sprachgeschichte*. Max Niemeyer, Tübingen.
- Popper, K., 1968. *The logic of scientific discovery*. Harper, New York.
- Popper, K., Eccles, J., 1977. *The self and its brain*. Springer International, Berlin.
- Reintges, C., 2009. Spontaneous syntactic change. In: Crisma & Longobardi (Eds.), pp. 41–57.
- Roberts, I., 2000. Caricaturing dissent. *Natural Language & Linguistic Theory* 18, 849–857.
- Roberts, I., 2007. *Diachronic syntax*. Oxford University Press, Oxford.
- Roberts, I., Roussou, A., 2003. *Syntactic change: a Minimalist approach to grammaticalization*. Cambridge University Press, Cambridge.
- Sundquist, J. D., 2010. Variation, continuity and contact in Middle Norwegian and Middle Low German. In: Breitbarth et al. (Eds.), pp. 145–168.
- Waltereit, R., Detges, U., 2008. Syntactic change from within and from without syntax: a usage-based analysis. In: Detges & Waltereit (Eds.), pp. 13–30.
- Willis, D., 1998. *Syntactic change in Welsh: a study of the loss of verb-second*. Clarendon, Oxford.
- Yang, C., 2002. *Knowledge and learning in natural language*. Oxford University Press, Oxford.
- Yang, C., 2004. Universal grammar, statistics or both? *Trends in Cognitive Sciences* 8, 451–456.