

This PDF includes a chapter from the following book:

Born to Parse

How Children Select Their Languages

© 2020 Massachusetts Institute of Technology

License Terms:

Made available under a Creative Commons
Attribution-NonCommercial-NoDerivatives 4.0
International Public License

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

OA Funding Provided By:

The open access edition of this book was made possible by generous funding from Arcadia—a charitable fund of Lisbet Rausing and Peter Baldwin.

The title-level DOI for this work is:

[doi:10.7551/mitpress/12799.001.0001](https://doi.org/10.7551/mitpress/12799.001.0001)

5 Population Biology: The Spread of New Variable Properties

5.1 Discontinuities and Abstractions

If syntacticians walk the four hundred miles from Berlin to Amsterdam, covering ten miles a day, it is said that they will hear no noticeable difference in the language of local people they meet at breakfast and at dinner.¹ Not even on the day when they cross the border from Germany into the Netherlands. The paradox is that the German of Berlin is indeed different from the Dutch of Amsterdam.

As in space, so in time. The messages that my daughters send are similar to the letters my mother used to send. The syntax is the same and the vocabulary close to identical, though they might use a few different words with different meanings. But a similar paradox applies: local differences are slight, but if we look over longer time spans, we see big differences across the language of Geoffrey Chaucer, William Shakespeare, Jane Austen, and Toni Morrison.

In many contexts, between neighboring towns and villages or between generations of a family, language seems to be stable and its transmission frequently seems close to perfect. Change is often gradual to the point of being imperceptible, but when we use a longer lens, we see major discontinuities.

There are many understandings of this paradox, depending on one's broad view of what drives syntactic change and acquisition. Nineteenth-century neogrammarians developed a theory of sound change but nothing parallel for syntax, morphology, or other aspects of language. This was their legacy for American structuralists, who also limited their work to what we now call phonology and to narrowly descriptive work in syntax and morphology.

A typical, linguistics-textbook treatment from the 1950s depicts language change as change in sounds and, crucially for our discussion here, as necessarily gradual: "we described sound change as a gradual change in habits of articulation and hearing, taking place constantly, but so slowly that no single individual would ever be aware that he might be passing on a manner of pronunciation different from that which he acquired as a child. This gradualness is extremely important" (Hockett 1958: 439). Hockett goes on, "when a person speaks, he aims his articulatory motions more or less accurately at one after another of a set of bull's-eyes, the allophones of the language" (p. 440). Speakers are "quite sloppy in [their] aims most of the time" and over time may hit different targets, hence language change. Thus, for many in the 1950s, language change was gradual change in sound production.

Kiparsky 1968: 175 describes this view as tantamount to seeing a language as a "gradual and imperceptibly changing object which smoothly floats through time and space," changing, for example, from Old, to Middle, to Early Modern, and then to Present-Day English, with various gradations in between but no discontinuities or major disruptions. That view is still often expressed, as we shall see.

Kiparsky and his colleagues, on the other hand, were early generativists dealing with change. They worked with abstract grammars encompassing syntax and morphology alongside sounds and viewed them as changing when children encountered new ambient language. An abstract grammar was a system with its own struc-

tures and computational operations, a biological object selected by children and represented in people's brains. For Kiparsky, "the transmission of language is discontinuous, and a language is recreated by each child on the basis of the speech data it hears." If one abstract structure in a generative system changes, that typically has multiple consequences for the structures and expressions generated by the system, leading to discontinuities and possibilities for large-scale changes.

Although change is often gradual to the point of being imperceptible, the major discontinuities that arise stem from the abstractness of the structures. Lightfoot 1979 and 2017c reiterates Kiparsky's view that grammars are invented or selected by children, not transmitted wholesale to the next generation, and summarizes arguments for some major and now well-understood discontinuities. I show how the paradox of the syntactician walking through imperceptible variations from Berlin to Amsterdam and ending up in a different language might be explained by a nonstandard approach to language acquisition that employs a clear distinction between sociologically defined external language and the internal languages of minds/brains. I argue that children discover and select the abstract structures of their internal languages. New phenomena in the input may trigger a single change at the abstract level, yielding a new structure that serves to generate many new phenomena that enter the language at the same time.

This approach to language change anticipates neo-Darwinian biologists' appeal to punctuated equilibrium, in turn based on evolutionary biologist Ernst Mayr's model of geographic speciation (Mayr 1942; Eldridge & Gould 1972), and their rejection of Darwin's gradualism. In many disciplines, this focused attention on structural shifts, known as "catastrophes" (Thom 1972)² or "phase transitions" at different stages of investigation within the successive frameworks of catastrophe theory in the 1950s and 1960s; chaos theory (Gleick 1987) and synergetics (Haken 1984) in the

1970s and '80s; and complexity science more recently (Casti 1994; Kauffman 1995; Prigogine & Stengers 1997).

Linguists have identified such saltations and understood them in terms of children selecting new systems when exposed to new ambient language. Sometimes several phenomena change simultaneously, a catastrophe or phase transition, and one can explain the simultaneity by arguing that there is a single change at the abstract level of the internal system from which the new phenomena follow. The task then is to show how the new system might have been discovered and selected by children; how new ambient language (i.e., new external language) triggers new internal systems. Rich and deep explanations have been developed for some syntactic changes, as outlined in §2.4–§2.6 and §3.3–§3.5. These explanations were deepened by a conceptual shift in the 1980s, which might be characterized as the great individualization.

5.2 Individualism

Noam Chomsky famously defined his field as “concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance” (1965: 3). Under this idealization, linguists wrote grammars for complete languages like Finnish, English, Japanese, and so on, capturing the competence of idealized speaker-listeners.

However, a subsequent work, Chomsky 1986, elaborated this competence–performance idealization and introduced a distinction between external language and internal languages, very different kinds of things. As discussed in §2.2, E-language is a mass-sociological concept, a group phenomenon, external language out

there, the kind of unanalyzed thing that anybody might hear, and it includes the PLD that trigger new systems. I-languages, on the other hand, are internal, individual systems that emerge in children according to the dictates of the inherent language capacity and to the demands of the ambient E-language to which they are exposed and which they seek to understand and analyze, that is, to parse. I-languages are properties of individual minds/brains and consist of abstract structures. Individuals acquire an individual, private I-language, creating their particular form of English, perhaps the language of a seventy-five-year-old woman in Padstow in northern Cornwall, and not the external English language as a whole or even the English of her community.

Chomsky now echoed Wilhelm von Humboldt, who distinguished the language of individual citizens from the language of a nation ([1836] 1971), and Hermann Paul, who asserted that “die reelle Sprache nur im Individuum existiert” (real language exists only in individuals; 1877: 325) and later that “wir müssen eigentlich so viele Sprachen unterscheiden als es Individuen gibt” (we must in fact distinguish as many languages as there are individuals; 1880: 31).

Put differently, “English” is not recursively enumerable. There is no system that will generate the sentences of English, partly because of internal contradictions. For example, an expression like *Kim might could read this* is an expression of English in Arkansas but not in New York or Cornwall. *Gwain ee t'Exeter* ‘is he going to Exeter?’ is said in Cornwall, perhaps by our seventy-five-year-old Padstow woman, but not in New York or Arkansas. Under this view, the English language has no more reality than the French liver, English irony, or the Scottish love of whisky; such things do not exist except as abstract idealizations.

Similarly, and relevantly for the catastrophe/phase transition to be discussed in §5.4, there is no recursive device that generates the set of expressions found in surviving English texts from the

thirteenth and fourteenth centuries. Rather there is a *set* of I-languages that generates what we see in the texts. Thus, we study language change in part as the spread of new I-languages, using the methods of population biology (see §5.9). There is no biologically coherent notion of English, certainly not as an object being transmitted steadily from one generation of language users to another. If languages are not transmitted, there are many damaging consequences for traditional ways of thinking about language change, but that is a topic for other days.

Chomsky made his move toward I-languages at about the same time that Tony Kroch introduced his idea of competing, coexisting grammars: individuals often have more than one internal system and may use different systems at different times in a kind of internal multiglossia (1989; see also Yang & Roeper 2011). If individuals may use more than one I-language, then we may go beyond Paul and argue that there are many more I-languages than there are individuals. Bear in mind that, under the approach we have adopted here, a person's individual I-language is the means by which she parses what she hears, so she may need multiple I-languages to deal with the diversity of the ambient language. It may not be possible to parse what one hears from visiting relatives of different ages and from different countries using a single I-language.

These ideas have had profound consequences for syntactic analyses and, in particular, for thinking about syntactic change, consequences that have still not been fully thought through. They enable us to gain new understanding of catastrophes or phase transitions, when many phenomena change at the same time, or of domino effects, when changes occur in rapid succession. The ideas of E-language and I-languages suffice to account for language acquisition, without the conventional, sociologically defined notion of English or Estonian. With these notions, linguists do not write grammars for whole languages but for I-languages that capture an individual's linguistic capacity. One views children as selecting

their private I-language on exposure to E-language. Children parse the E-language they hear and acquire the categories and structures needed to understand what they hear. Thus they discover and select the elements of their I-language (Lightfoot 2017b). A person's I-language is the aggregation of the structures permitted by UG and those required to understand what she hears (i.e., those already discovered and selected).

In §5.3 I will sketch some views about the relationship between major discontinuities in I-languages, catastrophes or phase transitions, and what writers regard as unusual events. In §5.4 I will outline the I-language–E-language distinction in greater detail and sketch how our discovery-and-invention/selection approach to acquisition enables us to understand the emergence of new I-languages with multiple new variable properties, specifically in the remarkable development of Middle English. In §5.5–§5.6, I will outline again our particular discovery approach to acquisition and consider through that lens the central paradox that we defined at the beginning of this chapter. In §5.7, we will work through some details of discontinuities first discussed in chapter 2, by considering how children were driven to the discontinuities revealed through the new internal systems, the new I-languages that children grew. In §5.8, we will think about the spread of new I-languages, and in §5.9, we will use the methods of population biology to explore phase transitions in language and will discuss a computer simulation.

5.3 Discontinuities and Unusual Events

The idea that language transmission generally takes place more or less perfectly and that children acquire the same language capacity as their neighbors and parents (unless something unusual happens) is still with us. For example, Jürgen Meisel has argued (2011) that change takes place only in multilingual contexts, which he

takes to involve sociologically distinct languages like German and Turkish. Discontinuities are viewed as unusual events provoked by particular social contexts, as when people are exposed to multiple languages, in the conventional sense of the term. So in recent generations, the multilingualism of Turks and Germans in Germany has had effects on the two languages.

However, multilingualism in Meisel's sense does not explain how languages like Icelandic, which have been isolated for a long time, nonetheless undergo changes in their syntactic structures, including major shifts in word order. Furthermore, if one works with conventional definitions of languages as properties of groups, one underestimates the variation and multilingualism found everywhere (Lightfoot 2011). There is significant variation among people within the same speech community, indicating individuals have their own private systems. In §5.5, we will reject the conventional notion of a socially defined language as the locus of change and distinguish internal and external languages, arguing that real change is best defined at the level of individual language users (Lightfoot 1993).

A similar predilection for stasis in normal times, with change only as a consequence of unusual events, comes from recent ideas about "inertia," developed quite differently in Longobardi 2001 and Keenan 2002 (see Roberts 2017 for good discussion of the differences). The central idea is that "things stay as they are unless acted upon by an outside force or decay" (Keenan 2002: 327) or that "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 changes and semantic changes, including the appearance/disappearance of whole lexical items) or, recursively, of other syntactic changes" (Longobardi 2001: 278). These ideas were developed in part as a reaction against the notion that changes may be internally motivated by "UG biases" and cyclical forces (Roberts & Roussou 2003; Van Gelderen 2011). Inertia approaches require an external cause for change, but the external

cause might be the result of a prior change in internal grammars (for critical discussion, see Walkden 2012).

Another view is that sometimes there is “imperfect learning” (Kiparsky 1968; Trudgill 2002; Mitchener & Nowak 2004; Montrul 2008): children are exposed to the same linguistic experience as their parents, but their learning is imperfect and they converge on a different mature system, yielding imperfect transmission, and hence discontinuity. Imperfect learning may be a special event, but it is not triggered by new experiences and can take place at any time, yielding discontinuities randomly.

5.4 The Case of Anglicized Norse

Emonds and Faarlund (2014) offer a radical challenge to the philologists’ conception of English progressing gradually and imperceptibly from one stage to another. The authors, henceforth E&F, postulate that so-called early Middle English—spoken and, crucially for historical linguists, *written* in the East Midlands in the twelfth and thirteenth centuries—reflects the results of intensive and extensive contact and represents a new language, which they call Anglicized Norse. It has many features of Scandinavian syntax alongside West Germanic words and phonological structures corresponding to Old English antecedents. They argue that Anglicized Norse syntax essentially replaced that of Old English. E&F build on work by philologists and syntacticians who argue for analyses that E&F construe now as aspects of a more comprehensive phase transition, and they muster considerable evidence for their analysis. However, their work has had, it is fair to say, a hostile reception from many historical linguists, notably in an issue of *Language Dynamics and Change* devoted to discussion of their book by several authors (e.g., Holmberg 2016; McWhorter 2016; Trudgill 2016) and in a review article by Kristin Bech and George Walkden (2015). For a more sympathetic view, see Lightfoot 2016,³

where I claim that the analysis is being rejected and discarded prematurely.

New I-languages emerge when the ambient E-language experienced by children changes. E&F offer an intriguing sociopolitical history of language used in medieval England, addressing several matters that have been raised by traditional historians of English. They discuss how England was subjugated by Scandinavians for two hundred years and then both the English and Scandinavians were oppressed mercilessly by the Normans. By 1100, all property of any note was in the hands of the Normans and “two previously separate peoples became united in servitude” (p. 41). “The miserable circumstances gave rise to a complete fusion of two previously separate populations, speakers of Old English and speakers of Scandinavian” (p. 43). They intermarried and there was much bilingualism (O’Neil 1978). This, and not when the Scandinavians first arrived and constituted the ruling class, is when we begin to observe significant Scandinavian influence on the native written language. Despite both Germanic populations being dispossessed, the Scandinavians predominated in trade, in agriculture, and in leading the opposition to the French. The Scandinavians settled permanently in the East Midlands and North and seem to have enjoyed notably higher economic status than the native English. A plausible sociopolitical history of language in England enables us to understand better how eventually the dominant features of external language, both spoken and written, came to be Norse for many people. It explains why children came to acquire Norse syntax, given that their ambient E-language had changed and now incorporated many aspects of Norse syntax. And if they acquired elements of Norse syntax, we understand why there was a wholesale introduction of new constructions in the written language, as Old English became more and more restricted to impoverished and illiterate peasants and eventually died out in a kind of “language replacement,” discussed in Campbell 2015.

The external history that E&F provide enables us to understand how the new I-languages that they postulate might have been discovered and selected by children, but the substance of their case lies in the linguistic analyses they outline. They tease apart syntactic characteristics of West Germanic (Dutch, Frisian, High and Low German, and their later offshoots Yiddish and Afrikaans) and North Germanic (mainland Scandinavian: Norse, Norwegian, Danish, Swedish), all well-analyzed languages, and they argue that Old English, to a large extent, has the characteristics of the former but Middle English the latter. For example, North Germanic has underlying head-initial VPs, West Germanic “at least partly” head-final VPs; North Germanic has infinitives with a *to* free morpheme, while West Germanic uses inflection; North Germanic has subject raising but West Germanic does not; restrictive relatives are introduced by invariant morphemes in North Germanic but by pronouns marked for case in West Germanic. E&F examine twenty such structures, which they call “parameters,” and argue that early Middle English speakers began to set them in the North Germanic fashion. For all twenty parameters, they offer a plausible demonstration that Old English speakers set them in the West Germanic fashion and early Middle English speakers in the North Germanic fashion. Furthermore, Holmberg 2016 notes strikingly about E&F that “as for syntactic features that Middle English shares with Old English that are not shared with Norse, they don’t find any!” Holmberg also notes that there seem to be other ways in which Middle English adopted Scandinavian characteristics, beyond what E&F consider, for example the deletion of complementizers and the avoidance of *that*–trace violations.

Whatever the details of each of the twenty structures, there was clearly a major discontinuity between Old and Middle English. The commentators are clearly strongly committed to the idea that languages change only slowly. But perhaps more will emerge in future discussions. For example McWhorter 2016 suggests that E&F gloss

over factors that indicate that Middle English is too different from Norse for their hypothesis to be sustained.

There is much to be said about all twenty features, but it is noteworthy how little is said about them in the commentaries offered so far and how things that are said often turn out to be misconceived—see below. Peter Trudgill (2016) is persuaded that “Emonds and Faarlund have brilliantly demonstrated that the syntax of my native language owes a great deal to the syntax of [Old Norse]—and very much more than has generally been thought.” Noting E&F’s “deeply impressive achievement,” he laments “the generativist mindset” of the authors, but aspects of the generativist mindset not called on by E&F in fact strengthen their case, namely the E-language–I-language distinction and matters of language acquisition.

E&F offer excellent discussion (pp. 84–93) of the very unusual property of preposition stranding, absent in most Indo-European languages including Old English and West Germanic (except Frisian and Dutch under very special circumstances) but found in the early and modern mainland-Scandinavian languages (compare also Walkden 2017 and Thoms 2019). E&F also offer nuanced discussion of the change from head-final to head-initial VPs, recognizing work showing the new V–DP order occurring sometimes in Old English texts (e.g., Pintzuk 2002).

It is important to note that, whatever the descriptive success with the twenty properties discussed, E&F attain a remarkable level of *explanatory* adequacy by postulating that the language of Norse speakers played a major role in triggering the syntactic structures of the English speech community during the early Middle English period (the twelfth and thirteenth centuries). For example, E&F discuss the verb-second properties of Old English, which show the complicating property of verb-third when the subject is pronominal and to the left of the finite verb (Van Kemenade 1987): [ælc yfel] *he mæg don*, ‘each evil, he **can** do’. Whatever this special property

of Old English is, it is absent from Norse and therefore, given E&F's central hypothesis, it is predictably absent from Anglicized Norse and early Middle English. By claiming that it was Norse syntactic systems that emerged, E&F *predict* that all the twenty relevant properties should have emerged in the first Middle English texts and that they should not have emerged in piecemeal fashion: all the new phenomena involve "changes" in the direction of North Germanic parameter settings. On another hand, for Bech and Walkden (2015), for example, this consistency is either accidental or not real.

E&F write, "The Old/Middle English break very much concerns the structure of the language itself; it is very little connected with how English was used or how it was perceived" (p. 28). "When English began to be written after the [Norman] Conquest, the new characteristics were clearly in the ascendant, most strongly in the former Danelaw ... while many aspects of Old English (as well as most of its vocabulary) had disappeared or been reduced to remnant percentages, especially in the South and Southwest" (p. 29).

5.5 Discovering and Selecting Elements of I-Languages

It is common within generative circles to view children as comparing the success of grammars in generating structures that match those elicited in the ambient language. An alternative stance is to embrace the antipositivist arguments of Chomsky 1975, that there is no procedure to guarantee that scientists discover or decide on correct theories; the best they can do is to compare theories and decide which one gives a superior account of some predetermined data, is learnable, achieves greater depth of explanation, and so on. A third view is that children are not constructing theories but subconsciously acquire a system that characterizes their linguistic capacity. We have argued here that they may be following a kind of discovery procedure but are limited to having structures generated by the Project and Merge operations allowed by UG.

Before we come to the discovery procedure itself, let us go back to the idea that the reason linguists have never been able to provide a clear definition of a language as distinct from a dialect is because the sentences of, say, English do not constitute a recursively enumerable set. To repeat, if one asks if *She might could do it* is a sentence of English, the answer is that it is a sentence of English in Alabama but not in Alaska. Therefore, there is no English such that it is something acquired by children growing up in Alabama, Alaska, and Alice Springs. Rather, children acquire different capacities, different I-languages, and one needs something more granulated to capture this. The sociological notion of English seems to be not relevant for an account of people's biological language capacity (but we return to this matter in §5.9).

We noted in §5.2 that Wilhelm von Humboldt distinguished languages as the creations of nations from languages as the creations of individuals: language "is not a mere external vehicle, designed to sustain social intercourse, but an indispensable factor for the development of human intellectual powers... While languages are... creations of nations, they still remain personal and independent creations of individuals" ([1836] 1971: 5, 22). We have observed in several places that Chomsky 1986 follows in this vein and distinguishes external language and internal, individual languages. E-language refers to language out there in the world, the kind of thing that a child might be exposed to, an amorphous mass. I-language, on the other hand, refers to a biological system that grows in a child's mind/brain in the first few years of life and characterizes that individual's linguistic capacity. It consists of structures, categories, morphemes, phonemes, and a set of computational operations that copy the items, delete them, assign indices to them, and so on. One's I-language is a private object that permits communication with certain other speakers, although it is not necessarily identical to the I-languages of those speakers.

These ideas of E-language and I-languages suffice for the purposes of accounting for language acquisition, and we do not need the conventional, sociologically defined notion of English or Estonian. With these ideas, one can view children as selecting their I-language on exposure to E-language. Rather than evaluating systems against a set of data, children can be viewed as paying no attention to what any I-language or grammar generates but instead growing an I-language by discovering and selecting its elements (Lightfoot 2006a). This is a procedure whereby children exploit the UG toolbox they have at birth and parse the E-language they hear, discover the categories and structures needed to understand what they hear, and thereby select the elements of their I-language. E-language and I-languages are quite different kinds of entities but they are related: elements of I-languages are triggered by phenomena of E-language, and that is the discovery mechanism.

UG provides children with the set of structures that they might need in order to understand and parse the E-language that they are exposed to; these are the structures that robust E-language elicits in them. Children are born to parse and, as noted in §3.1, after they know that *cat* is a noun referring to a domestic feline and *sit* is an intransitive verb, they may hear an expression *The cat sat on the mat* and recognize that it contains a DP consisting of a determiner *the* and a noun *cat* and a VP containing an inflected verb *sat* followed by a PP *on the mat*. The child makes use of the structures needed to parse what is heard, and once a structure is used it is incorporated into the emerging I-language. In this way a child accumulates the structures of their I-language, which are required to parse the ambient E-language; children select elements for their I-language, the basic structures, in piecemeal fashion.

At no stage do children calculate what their current I-language can generate. Rather, they simply accumulate the necessary structures. Furthermore, if UG makes available a thousand possible

structures for children to draw from, that raises no intractable feasibility problems comparable to those facing a child evaluating the generative capacity of grammars with thirty possible parameter settings, checking the grammars against what has been heard. There are no elaborate calculations. Children developing some form of English I-language learn without apparent difficulty irregular past-tense and plural forms for a few hundred verbs and nouns. Learning that there is a structure $_{VP}[V + \text{Infl PP}]$ seems to be broadly a similar kind of learning, although much remains to be said (for detailed discussion, see Fodor 1998a; Dresher 1999; Lightfoot 1999).

5.6 Individual Languages versus Big Data

E&F achieve considerable descriptive and explanatory success in their account of the emergence of “Middle English,” now equipped with scare quotes. However, their account is troublesome for people who believe that language changes only gradually and imperceptibly. In contrast, we can understand why and how the changes E&F describe should have happened if we think in terms of I-languages being selected by individuals. We have emphasized that I-languages exist for people and not for languages; there is no grammar of English in any biological sense. Therefore, rather than thinking of “language change,” we need to think less grossly, in terms of the spread of new I-languages. Humboldt, Paul, and modern work take a biological view of languages, as opposed to a social view. At a minimum, different questions arise under each view and the same questions take on quite different complexions; this is worth some reflection.

The overwhelmingly most common view among historical linguists is that language change is gradual. That view seems to drive E&F’s antagonists in *Language Dynamics and Change*. However, much depends on the units of analysis, the kind of lens used. Lan-

guages seen as social entities can be seen as changing gradually, but I-languages emerge abruptly in an individual speaker.

Fries 1940 looks through a wide-angle lens and finds that Old English around the year 1000 shows object–verb order 53 percent of the time. That order is gradually replaced by verb–object order, reducing to 2 percent by the year 1500. Fries provides one set of statistics for each century but offers no analysis. His counts ignore the distinction between matrix and embedded clauses, where word orders are different, and he has no explanation for the fact that the finite verb often appears in second position in simple clauses. If one makes such distinctions, one can count more productively and see that Old English I-languages had object–verb order underlyingly, with a system yielding subjects in first and third position, and objects “extraposed” to the right (but the widely imitated analysis of Dutch and German, moving finite verbs to a higher C position in matrix clauses, which Van Kemenade 1987 applies to Old English, is not supported). Consequently, we find object–verb order uniformly in embedded clauses, but only variably in matrix clauses. In fact, at least two distinct changes took place in I-languages at different times: object–verb order was replaced by verb–object, and the operation moving objects to the right of the verb was lost (Haerberli 2002a,b).

If units of analysis are as gross as Fries’s, change will look gradual. But one must be wary of “big data,” often gathered these days through digitized corpora that do not make the E-language–I-language distinction that is essential to our view. In gathering data from a certain period, say fifteenth-century England, one must resist the temptation to assume that there must be a device that will generate the entire collection of data. One must be ready to distinguish abstract structures generated by new I-languages and the raw data of E-language that might trigger those new I-languages.

At the other end of the scale, if we use a telescopic lens, the speech of no two people is identical, change is everywhere, all is

in flux, languages are constantly changing in piecemeal, gradual, and minor fashion; again we see constant, gradual change. Initial experiences are never entirely the same for two speakers, and they may differ in minor and insignificant ways. A particular construction type might become more frequent, perhaps as a result of taking on some expressive function. This does not reflect the properties of an I-language itself but the way in which I-languages are used. Such changes in frequency do not reflect a change in I-languages, but they do entail a change in the external language for the next generation of speakers, therefore for the PLD triggering the next I-languages.

Not only may E-language change gradually, the very nature of language acquisition ensures a kind of gradualness in that children experience the speech of their parents, older siblings, and other household members. This gradualness works against major discontinuities in the class of expressions and their associated meanings. For example, one generally does not find an I-language that yields more or less uniform object–verb order being replaced abruptly by one that yields uniform verb–object order. Even so, one does find significant discontinuities, as E&F have shown dramatically, especially in contexts where the output of a parent’s native I-language does not contribute as significantly to a child’s experience as the I-languages of children who have already acquired elements of Norse syntax. To be sure, there is no reason to believe that there is any formal relationship between the I-languages of parents and children. I-languages are created afresh by every individual and may differ in form from those of their parents, perhaps radically, within the limits of UG. For discussion, see Lightfoot 1999: chap. 4, where I-language changes are viewed as Thom-style catastrophes or phase transitions; there may be gradual change in the temperature of water until there are structural changes only at the critical points where it becomes a solid or gas (Thom 1972; Casti 1994). Now see Haerberli and Ihsane (to appear) for interesting new evidence corroborating one of the best known cases of reanalysis in the history of

English, that of the modal verbs (§2.4).⁴ Similarly, E-language may change gradually and trigger a new I-language at a certain point (see Westergaard 2017 for discussion of the gradualness of change from a range of theoretical perspectives).

Whether differences between I-languages are small-scale or large-scale, these differences do not have temporal properties, and changes in I-languages cannot be “gradual”: a person either has a particular I-language or another one. Apparent gradualness does not reflect an I-language property but a mirage, conjured by a failure to distinguish independent change events in E-language from changes of I-language. E-language may differ in ways that do not trigger a new I-language. However, a natural way for linguists to think of catastrophic changes is to envisage E-language crossing a threshold, which triggers a different I-language system. The relatively few variable properties given by UG (the abstract structures) constitutes the set of “fixed-point attractors,” familiar from chaos theory (see Kauffman 1995) defining the nature of possible changes. An I-language either has some property or does not.

5.7 Back to Discontinuities

Equipped now with these ideas of E-language and I-languages, seeing language acquisition as reflecting a drive to parse E-language and to discover and select the elements of I-languages, one can understand historical discontinuities as new I-languages triggered by new E-language, that is, by new PLD. Children parse the ambient E-language and select the structures of the emerging I-languages that coexist in a single speech community and in a single speaker. Discontinuities do not constitute a special event. Rather they are the normal state of affairs confronting any child. There is no “imperfect learning” or “imperfect transmission,” just different transmission. New E-language triggers new I-languages. New E-language is the initial locus of change, and furthermore, we can link particular

aspects of E-language to new I-languages. Nothing is transmitted. Elements of I-languages are selected afresh in each generation and in each individual.

If work on language change provides insight into linking particular aspects of E-language to new I-languages, that constitutes a major contribution to our understanding of children's language acquisition, not achieved so far in experimental work on children. Furthermore, as in so many other domains, observing how a thing changes often reveals properties of that thing. To see how that works, let us elaborate on two changes in the I-languages of English speakers, discussed preliminarily in chapter 2. In §2.4 we noted that by Early Modern English, but beginning earlier, I-languages had *can*, *could*, *may*, *might*, *must*, *shall*, *should*, *will*, and *would* categorized or parsed as Infl items, whereas they had been verbs for earlier speakers. As a result, they came to have a more restricted syntactic distribution, ceasing to occur in various contexts. There is good reason to believe that this change in I-languages was triggered by a prior change in E-language.

Old English I-languages showed many inflections, indicating the tense, person, number, and conjugation type of verbs and the number, case, and declensional class of nouns. While regular Modern English verbs show up in four forms (*refuse*, *refuses*, *refused*, *refusing*), Old English verbs had over a hundred forms, as observed in §2.4. However, all of that was vastly simplified over the course of Middle English. The morphological distinctions were eliminated first in the north of England (manifested first in the Lindisfarne Gospels) and later in London and the south, arguably an effect of widespread English–Scandinavian bilingualism (O'Neil 1978). Individuals with English and Scandinavian I-languages had a rich morphological system, similar to each other but different and not learnable as a single system.

The only aspects of present-tense verb morphology to survive the great Middle English simplification were the third-person singular

ending *-s/-eþ* and the second-person singular ending *-st*. It is quite unclear why those elements of present-tense verb morphology should have survived. However, the verbs that were to be recategorized as Infl elements had been members of the preterite-present class, which had what were usually past-tense endings for forms of the present tense, a phenomenon that also occurs in Latin verbs like *coepi* ‘begin’, *odi* ‘hate’ and Greek verbs like *oída* ‘know’ and *eoika* ‘seem’, which are present tense in meaning but perfect in form. The crucial fact about the morphology of the preterite-presents is that they never had the *-s/-eþ* third-person singular ending and therefore now lacked what was to become the single morphological property of present-tense verbs. Of the original preterite-presents, some dropped out of the language (e.g., *unna* ‘grant’, *benuga* ‘suffice’), others assimilated to regular verbal inflections with the third-person *-s* ending (*wita* ‘know’, *duga* ‘be of value’), and the remainder were recategorized as Infl items.

Furthermore, with the loss of subjunctive endings as part of the morphological impoverishment, another defining property of verbs came to be that *-d* forms indicated past tense. However, the *-d* forms of the items to be recategorized, *could*, *might*, *must*, *would*, *should*, rarely indicated past tense (**She might lift eighty kilos until yesterday*) but rather retained “subjunctive meaning.”

One of the tasks of a child developing his/her I-language is to identify the words and the categories to which they belong, and this is done on the basis of formal and distributional properties and is the role of parsing. After the simplification of complex verb morphology, having the third-person singular *-s* ending became a defining property of verbs, but *can*, *could*, *may*, and so on did not have it. Verbs had past tenses in *-d*, but *might*, *would*, *should*, and *must* almost never carried past-time meaning and *could* only rarely did (*She could lift 75 kilos until she turned thirty*). Therefore, after the great simplification of morphology, the old preterite-presents lacked what had become the defining formal properties of verbs. They

could no longer be verbs and were assigned to the only plausible category that could be immediately followed by a VP: Infl or T, depending on one's theory of functional categories. Hence the newly restricted syntactic distribution. If *can*, *may*, and so on were instances of Infl or T, they could not occur to the right of an aspectual marker, a position restricted to lexical verbs (**She has could gone*; **Canning go, she left angrily*). They cannot occur with the infinitival *to* or with another modal verb, which are also instances of Infl/T (taken as a cover term for the relevant functional heads), because there can be only one such element in a clause (**I want to can leave*, **She might could read it*).

New E-language resulted from the contact between English and Scandinavian speakers in the north and the associated bilingualism, if O'Neil 1978 is right. That new E-language triggered new I-languages, with the old verbs *can*, *may*, *shall*, and so on recategorized as Infl items (as shown in Lightfoot 2017a). Hence the discontinuity and its explanation. There is no imperfect learning or imperfect transmission, just new E-language triggering a new I-language. The explanation for the change is local; it involves particular changes that took place in the ambient E-language at this time. The equivalent change has not taken place in closely related languages like Dutch and German or slightly more distant languages like French and Italian, because there were no comparable changes in E-language. Furthermore, we are not trapped into the circularity of claiming that the new I-language was triggered by exposure to the new data generated by the new I-language, a problem that confronts proponents of evaluation approaches to acquisition, as we saw in chapter 2.

Let us turn now to the second of the famous and well-understood reanalyses discussed in chapter 2 (§2.5), the loss of V-in-I structures, $_{\text{Infl}}\text{V}$, resulting from the movement of a verb to an Infl position. This loss is another change that has not taken place in closely related languages like Dutch and German. Under the discovery-

and-selection approach to acquisition, for such a structure to be discovered by children, they must *need* it in order to understand and parse utterances. For example, once Chaucer's son knew that *understand* was a verb and could occur in high positions in expressions like *Understands she this chapter?* or *She understands not this chapter*, he would know that the verb could occur in an Infl position and therefore also, for the first of these two utterances, in a C position. Such sentences in E-language expressed the $_{\text{Infl}}\text{V}$ structure. Sentences like the first also expressed the $_{\text{C}}[\text{Infl}\text{V}]$ structure. Similarly, Thomas More, for whom *can* and *may* and so on were verbs, would also understand *She cannot understand this chapter* as expressing the $_{\text{Infl}}\text{V}$ structure, because the verb *can* occurs to left of the negative *not* and therefore must have moved to the higher Infl position. Likewise, a French-speaking child hears *Elle comprend pas ce chapitre* 'She understands not this chapter' as expressing the $_{\text{Infl}}\text{V}$ structure. As does a Dutch child hearing *Begrijpt zij dit hoofdstuk?* 'Understands she this chapter?' *Comprend* has moved to the left of the negative marker, to Infl, and *begrijpt* has moved to an Infl position, from which it can move to a higher C.

However, somebody younger than Thomas More, for whom *can*, *may*, and so on were not verbs but Infl items, would not parse *She cannot understand this chapter* as expressing the $_{\text{Infl}}\text{V}$ structure, because for them *can* would be an Infl item and therefore not a verb in the Infl position. Given that a clear majority of simple sentences contain a modal auxiliary (Leech 2003), this means that there was a large reduction in the expression of the $_{\text{Infl}}\text{V}$ structure.

Alongside the fact that sentences with a modal auxiliary no longer express the $_{\text{Infl}}\text{V}$ structure, expressions like *Understands she this chapter?* and *She understands not this chapter*, which reflected the $_{\text{Infl}}\text{V}$ structure, were giving way to forms with the periphrastic *do*: *Does she understand this chapter?* and *She doesn't understand this chapter*, where the $_{\text{Infl}}\text{V}$ structure is not expressed (*do* is not a verb moved to Infl; in these examples *understand* is the verb). These

forms entered the language in the late fifteenth century and spread from the southwest across the rest of the country. Ellegård 1954 provides a remarkably detailed account of that spread, and McWhorter 2009 argues that periphrastic *do* arose in the southwest under the influence of Cornish. (“Cornish’s auxiliary *do* presents a thoroughly plausible model for English’s periphrastic *do*” as a carrier of tense in negated and interrogative contexts [p. 164]. McWhorter goes on to show how this provides a good explanation for the spread of periphrastic *do* in Middle English.) The spread of periphrastic *do* further reduced the expression of the $_{\text{Infl}}\text{V}$ structure.

The evidence suggests that new E-language, stemming from the combination of the recategorized Infl items (an I-language change) and the new *do* forms, reduced the expression of the $_{\text{Infl}}\text{V}$ structure to below the threshold that would enable children to add it to the structures making up their mature I-languages. As a result of that change in I-languages, there were further changes in E-language, such that expressions like *Understands she this chapter?*, *She understands not this chapter*, and *She understood on Tuesday the chapter* no longer occurred.

In these two well-understood phase transitions, we see discontinuities that can be explained as responses to new E-language. In the case of the new Infl items, new E-language arose because of a prior change in I-languages, the dramatic simplification of morphological endings resulting from English–Scandinavian bilingualism. In the case of the loss of the operation moving verbs to a higher Infl position, E-language changed as an effect of the new Infl items just mentioned and of contact with Cornish, such that the old $_{\text{Infl}}\text{V}$ structure was no longer expressed sufficiently to be attained by children at this time. Nothing comparable has happened in the E-language that Dutch or German speakers experience, and therefore no comparable change in I-languages has occurred. In the case of the loss of the $_{\text{Infl}}\text{V}$ structure, however, unlike in the case of the recategorization of certain verbs, there are comparative data from

other languages. Heycock et al. 2012 reports the tail end of the loss of $_{\text{Inf}}\text{V}$ structures in Faroese, and Vikner 1995 argues that mainland Scandinavian has lost those structures.

We see changes in I-languages, discontinuities, that can be understood as responses to new E-language. There is nothing imperfect at work. Of course, no two children have exactly the same PLD; they hear different things. Nonetheless, despite variation in experience, children often attain the same mature I-language in terms of the set of known syntactic structures. Individual experiences may vary indefinitely, but I-languages show structural stability and vary only in limited ways. I-languages emerge in the usual manner on exposure to E-language that has changed in a critical way, so that the discontinuity occurs naturally enough. Neither E-language nor I-languages get transmitted, neither imperfectly nor perfectly. In particular, I-languages are not transmitted in any sense. Rather, they are invented afresh by young children on exposure to the PLD in E-language. They are not shaped by earlier I-languages except indirectly through E-language. As we noted, I-languages are not objects floating smoothly through time and space.

5.8 Gradual Change in Languages or Spread of I-Languages

We have emphasized that grammars, I-languages, exist for people and not for languages; there is no grammar of English in any biological sense. We noted antecedents for this view in nineteenth-century writings. Like Humboldt and Paul, we take a biological view of languages as opposed to a social view. At a minimum, different questions arise under the two views and the same questions take on quite different complexions; for discussion, see Lightfoot 1995, 1999: 79–82.

Everyday common sense suggests that, if there is little distortion, the patterns, processes, and structures of life do not change very much. Similarly with language. In addition, the overwhelmingly

most common view among historical linguists is that language change is gradual (cf. Westergaard 2017). But things depend on the units of analysis: languages seen as social entities change gradually, but I-languages change abruptly.

PLD may change gradually at first, and the very nature of language acquisition works against major discontinuities in the class of expressions and their associated meanings. For example, one does not find an I-language that yields more or less uniform object–verb order being replaced abruptly by one that yields uniform verb–object order. Even so, one does find significant discontinuities, especially in contexts where the output of a parent’s native I-language does not contribute significantly to a child’s PLD. To be sure, there is no reason to believe that there is any formal relationship between the I-languages of parents and children. I-languages are selected anew by every individual and may differ in form from that of their parents, perhaps radically, within the limits of UG. Whether differences between I-languages are small-scale or large-scale, they do not have temporal properties and cannot be gradual. Apparent gradualness of change is a mirage, conjured by a failure to distinguish independent change events, changes in E-language and changes in I-languages. PLD may differ in ways that do not trigger a new I-language. However, a natural way for linguists to think of significant changes is to consider different sets of PLD as sometimes crossing thresholds, which trigger different I-language systems. So the inventory of variable properties given by UG (the variable structures that may or may not occur in I-languages) constitutes the set of “fixed-point attractors,” that is, the set of I-languages, defining the set of possible changes.

The ideas of Kroch and his associates on competing grammars (Kroch 1989) factor into thinking about apparent gradualness of change and the diffusion of new I-languages. Paul thought that there were as many languages as individuals (1880: 31), but Kroch and colleagues (Kroch 1994; Kroch & Taylor 1997) maintain that there

are even more languages than individuals, because people operate with coexisting I-languages in a kind of internalized diglossia, indeed an internalized multiglossia. Their work enriches grammatical analyses by seeking to describe the variation we find within texts and across different writers, the spread of a grammatical change through a population. In postulating two or more coexisting I-languages in an individual, a researcher needs to show not only that the I-languages together account for the range of expressions used by that individual but also that the I-languages are all learnable under plausible assumptions about children's PLD. Multiglossic grammars are subject to exactly the same learnability demands as any other biological grammar.

This kind of diglossia provides an interesting approach to solving significant learnability problems. It offers a way to eliminate the unlearnable distinction between optional and obligatory operations (Lightfoot 1999: 92–101.). Chomsky 1995 argues that grammars do not permit optional operations. In that case, apparent optionality would be a function of coexisting I-languages. Instead of one I-language generating forms *a* and *b* optionally, one would argue that a person has access to multiple I-languages, one of which generates form *a*, another form *b*. The speaker has the option at any given time of using either of the I-languages. This move reduces the class of available grammars, eliminating those with optional operations (see also Biberauer & Richards 2006; Wallenberg 2013).

A consequence of this view is that Old English texts that only sometimes show verb-second phenomena cannot be explained by endowing Old English I-languages with a device for generating verb-second order optionally. Rather, there must have been competing I-languages, one generating verb-second order and the other not. Certain speakers have access to just one of these I-languages, other speakers have access to the other I-language, and others have access to both systems in an internalized diglossia. This turns out to be a productive analysis (Kroch & Santorini 2013).

On the view developed by Kroch, “change proceeds via competition between grammatically incompatible options which substitute for one another in usage” (1994: 180). One reason for believing that this view is along the right lines is that alternating forms cluster in their distribution; the clustering follows from how sets of I-languages unify the forms. We find not free variation but oscillation between two (or more) fixed points. This is reflected in the Constant-Rate Effect of Kroch 1989 and was noted in §2.5 in the context of Shakespeare using the old $_{\text{Infl}}V$ system as well as the new system without verb movement, alternating between the old and new systems within the same sentence.

Because structures are abstract, changing one element of structure or one categorization may entail a range of new surface phenomena. The Constant-Rate Effect requires that all surface phenomena reflecting the new I-language property have usage frequencies that change at the same rate, though not necessarily at the same time. This is easy to understand if one I-language is replaced over time by another, and if that change takes place in a winner-take-all competition between the two systems. We do not find complex arrays of linguistic data changing randomly. Instead, they tend to converge toward a relatively small number of patterns or attractors, in a kind of “antichaos” in the sense of Kauffman 1995. The points of variation defined by UG constitute the attractors and the two competing I-languages define the points of oscillation. Changes may occur when one I-language replaces another or when a new I-language is selected to coexist with another.

When we view an individual’s language capacity as characterized by one or more private, personal I-languages, then the spread of a new I-language across a speech community can be approached through the methods of population biology. An individual may be exposed to PLD that differ from what anybody else has been exposed to. This could happen because of population movements,

new patterns of bilingualism, or adult innovations by parents, caregivers, or teachers, or perhaps because the PLD are truncated in some way, not including earlier expressions or not including them with the same frequency as a generation earlier. If one individual selects a structure differently from others in the community, she is likely to produce different utterances. These new expressions, in turn, affect the linguistic environment, and our innovator will now be an agent of further change, reinforcing the PLD that might trigger another instance of her new I-language in younger siblings. As the younger siblings pick the same structures as their elder, so other people's PLD will differ, and a chain reaction is created. In this way a new I-language may spread in a way analogous to what has been observed in population genetics, replicating aspects of evolutionary change.

5.9 Population Dynamics and a Computational Model

Niyogi and Berwick 1995, 1997 present a computational model that analyzes change in this way and derives the trajectory of changes. Niyogi 2006 enriches the model. The model is based on a learning theory with three subcomponents: a theory of grammar, a learning algorithm by which a child generates grammars on exposure to data, and PLD. Niyogi and Berwick assume a population of child learners, a small number of whom fail to converge on preexisting grammars. After exposure to a finite amount of data, some of these children now converge on the preexisting grammar, but others attain a different I-language.

The next generation will therefore no longer be linguistically homogeneous. The third generation of children will hear sentences produced by the second—a different distribution—and they, in turn, will attain a different set of grammars. Over successive generations, the linguistic composition evolves as a dynamical system. (Niyogi & Berwick 1997: 2)

Emergence of a new I-language, in this simulation, is a logical consequence of specific assumptions about the theory of grammar, the learning algorithm, and the PLD. Interestingly, Niyogi and Berwick's model yields different time courses for different changes. A common trajectory is the familiar S-curve (Weinreich, Labov, & Herzog 1968; Kroch 1989): a change may begin gradually, pick up momentum, and proceed more rapidly, tailing off slowly before reaching completion. The success of Niyogi & Berwick is to build a dynamical system from a parameterized system and a memoryless learning algorithm. As a result, they *derive* the S-curve rather than building it into their model as a specific assumption. Further, the model allows that changing elements of the theory of grammar or of the learning algorithm may yield different trajectories, including trajectories other than an S-curve. Thus the model may be amended depending on how it matches the observed trajectory for specific changes in specific languages. This offers a new empirical demand for theories to meet, in addition to demands of learnability, coverage of data, explanatory depth, and so on: theories can be expected to provide the most accurate diachronic trajectories for identified changes.

Niyogi and Berwick 2009 recognizes that Chomsky 1965's idealization of speaker-hearers who use a homogeneous, social language led to a standard model of acquisition under which one PLD source developed into a single target grammar. Changing that idealization elicited population-based acquisition models. Niyogi & Berwick's model accurately captures how new I-languages progress through a community of speakers. This remarkable result clearly could not be replicated under a social definition of grammars, which denies the usefulness of individual, biological grammars. There may be slowness and gradualness in the spread of a change through a population, but changes in I-languages are rapid and abrupt at the individual level; familiar S-curves generally arise as a function of averaging across groups. The rapidity of change

follows from an extension of the intuition behind Mark Aronoff's Blocking Constraint, which limits coexisting forms to those that are functionally distinct (Aronoff 1976).

This all strongly suggests that structural changes are rapid and abrupt at the individual level and often spread through a population rapidly. The speed of the spread depends on nongrammatical factors relating to social cohesion, facility of communication among different groups, and so on.

5.10 Conclusion

The overwhelming consensus among historical linguists is that languages change gradually. It is granted that there are significant discontinuities that happen from time to time, but they are allegedly due to unusual events. We have outlined a different approach here in terms of a distinction between amorphous and constantly shifting E-language, on the one hand (constantly in flux and never experienced the same way by any two people), and biological I-languages on the other hand (represented in the minds/brains of individuals, recursive systems that characterize the individual's language capacity). We view an I-language as emerging in a child as its elements are expressed in the ambient E-language and are selected by the child. I-languages are internal, individual entities, and speakers typically operate with more than one I-language. Different I-languages may be selected when children are exposed to different E-language.

Construing a person's language capacity as an individual, private matter, we can understand how different linguistic experience may trigger a different internal system, which may then spread through a speech community in ways that can be modeled through the methods of population biology. Indeed, the spread of many things, from political views to fashion to support of particular soccer teams, has been modeled in Niyogi's work. Under this view, discontinuities, new I-languages, are liable to emerge at any time and

can be understood as natural phenomena. Nothing is transmitted from one generation to another. Children develop an I-language when exposed to E-language that expresses certain structures. A child might develop a novel I-language, which may spread through a community. That is our understanding of “language change,” a derivative function that is best understood as an individual phenomenon that may affect the linguistic experience of others and lead to a shift in group behavior. By seeking to understand the emergence of new linguistic patterns through the acquisition of language systems by individuals, we can sometimes explain the new group behavior.

In this chapter we have emphasized the development of language systems as a property of individuals, sharing Hermann Paul’s view that languages belong to individuals. However, a paper by Hariharan Narayanan and Partha Niyogi (2013), written just before Niyogi’s untimely death, takes an intriguing approach, seeking to model how a group of linguistic agents might arrive at a shared communication system through local patterns of interaction, for example developing a shared vocabulary. If successful, this would derive the distinct group properties of, say, Navajo, Norwegian, and Nubian and thereby give them a basis in biological reality.

Understanding the emergence of new I-languages as a function of language acquisition leads us to different understanding of the apparent gradualness of change, recognizing discontinuities as a natural part of language history. I-languages are selected afresh by every individual on exposure to the ambient E-language, which is different for every individual. Languages do not get transmitted and language change is an epiphenomenon of individuals selecting their private I-languages.

The study of diachronic syntax is in its infancy. As it matures, through integration with the study of language acquisition, one would expect it to cast new light on lay ideas about languages and on their sociological character. And much more.