What forced syntax to emerge?

Dieter Wunderlich (Düsseldorf, Berlin)

1. Introduction. Chomsky (2005) distinguishes external merge (e-merge), which forms morphosyntactic units through local combination, from internal merge (i-merge), which forms syntactic chains through dislocation. The title of this paper refers to internal merge. I consider the conditions under which morpho-syntax (including head-marking morphology and adjunct syntax) emerged as different from those under which dislocation emerged. More precisely, I am going to argue that morpho-syntax was prior to dislocation syntax: that the former developed as soon as the linguistic capability of the modern human species came into existence, while the latter only developed in large communities involved in language contact, as it was typically the case at the end of the glacial period 10,000 or maximally 12,000 years ago.

The set of possible human languages, comprising both types of syntax, can be characterized by the notion of Universal Grammar (UG), which commonly is identified with the human-specific learning algorithm towards language (language acquisition device, LAD), which in turn can be regarded as the (genetically transferred) information to the brain how it has to process chunks of memorized linguistic input in order to find proper grammatical generalizations. Following Briscoe (2003), I assume that the LAD did not emerge in a single (and extremely unlikely) evolutionary step, but rather evolved incrementally in a number of independent steps. It is thus reasonable to assume that the several features considered by Hockett (1960, 1966) and others to be characteristic for human language (if contrasted with animal communication) have different origins (Wunderlich 2004). Concerning the evolution of the human genotype, the FOXP2 mutation (Lai at al. 2001, Enard et al. 2002a,b) seems to be the most recent change affecting human language, dated at about 160,000 (but not more than 200,000) years ago, which roughly coincides with the arrival of homo sapiens according to archeological evidence.

If it is true that dislocation syntax did not emerge before the end of the ice age, it should be regarded a cultural rather than a biological trait: it is not likely to be determined by the genotype because selection pressure is required to maintain a biological trait. Without such selection pressure for a period of more than 100,000 years, one would expect a trait to be whittled away by genetic drift. However, with such selection pressure, a newly emerged trait will continue to adapt, and will have been further refined by genetic assimilation (Briscoe 2003: 303). Therefore, one could at most accept the idea that dislocation syntax is a product of further genetic assimilation.

2. Evolutionary steps. Considering the prerequisites of human language, it is obvious that the evolution of upright moving freed the hands from being occupied by locomotion, and thus enabled the early hominids to begin with and improve manufacturing. According to the mirror system hypothesis (Rizzolatti & Arbib 1998), the neuronal mirror system for grasping, already known for monkeys (Rizzolatti et al. 1995), was gradually extended, which can explain why and how *intentionality* (the utterer intends communication to have a particular effect on the recipient) and *parity* (speaker and hearer roles are interchangeable), two of the most fundamental traits of human language, have evolved.

About the early man (homo ergaster/erectus) we know that he managed to maintain fire and to produce hand axes (1.5 - 1.4 mill. years ago). Both activities require some amount of intentional planning *beyond the here-and-now*, so that one can attribute to him the capability of *complex propositional thinking* (in the sense that a proposition is decomposed into predication and referential anchoring, and that second-order predication is possible). If it is true that hand axes served to be thrown (Calvin 1990), this new achievement in turn profited from an increasing size of the brain. Both *fast processing* and *mapping from hierarchical structure to temporal ordering*, two other fundamental features of human language, can be attributed to progresses in the timing of actions necessary for producing and using hand axes, that is, to sensomotoric skills that could have been adapted for language.

The advanced sensomotoric control of manual and facial expression, supported by further development of the neuronal mirror system, could have enabled the emergence of a first *gestural* language (Arbib's 2005 protosign, see also Corballis 2003), so that all the features independently developed for manual actions were transferred to language. Furthermore, the need of preparing mirror neuron assemblies in the brain can explain *the child's instinct of imitation*, which clearly distinguishes man from

chimpanzees and becomes the motor of language learning. Moreover, *language learning becomes the motor of further language evolution*. It is reasonable to assume that *cultural evolution*, enabled through iterated learning by members of subsequent generations on the basis of slightly modified inputs, already started in the time of homo erectus. Finally, a beginning with gestural language makes it plausible that iconic encodings (possible with gestures but not as easily with sounds) could be transformed stepwise to more abstract *symbols*. Signers are likely to have combined individual signs as much as they learned to combine hand movements in other actions. This suggests that some kind of *compositionality* (productive combination of signs) could have already played a role in protosign.

In a further step, possibly enforced by the FOXP2 mutation, the control of the vocal apparatus (including lips, tongue, velum and larynx) got improved, so that the *vocal-auditory modality* could take priority over the gestural-visual one. Protospeech could gradually free the hands from being necessarily involved in communication, which was an obvious advantage compared to a purely gestural system. As Studdert-Kennedy (2005) claims, and Oudeyer (2005) showed through simulation, both *discreteness* of articulatory features and the *duality of patterning* (minimal units do not bear meaning) emerge as automatic self-organizing consequences of random interactions among speakers-hearers. All more complex combinations then inherit discreteness from the minimal units.

This is the point where e-merge (or morpho-syntax) becomes interesting. In all sorts of combinations, some *basic asymmetry* is at work: one element functions as the head, whereas all others are non-heads. A vowel functions as the head of a syllable, and a stressed syllable as the head of a prosodic word. Complex predicates such as compounds and adjunction structures have one element as the head. The lexical inventory is partitioned into two complementary categorial types, verbs and nouns, with the verb as the head of a clause. Moreover, the information of lexical entries is asymmetric, so that the arguments are strictly ordered. Finally, the element in focus can be considered the head of a piece of discourse.

There is good reason to assume that *linguistic diversity* started just in the process of organizing articulatory features, and based on that, organizing the vocabulary, probably long before small groups of homo sapiens colonized all parts of the world (60 to 30,000 years ago). Of course, groups that have separated undergo separate changes, but random fluctuations in the very beginning are more likely to have produced the amount of phonological variation inherent to human languages. As Carstairs-McCarthy (2005) argues, *early morphology* could have started with a reinterpretation of allomorphic alternations produced by rapid speech. Here, again, a large amount of diversity could have evolved very early. Until today, morphophonology is perhaps the domain with the highest degree of cross-linguistic diversity.

In contrast, the amount of syntactic diversity seems to be smaller, and moreover it was often underestimated by syntacticians who stated particular syntactic universals, which later turned out to be rather generalizations concerning some striking linguistic families. Detailed typological studies revealed that many alleged universal notions or principles are in fact not characteristic for all human languages and thus are not universals. For instance, the notion of *grammatical subject* fails in the inverse type languages of the Algonquian family, and so does the notion of *grammatical subject-object asymmetry*. These languages also make the notion of *abstract case* doubtful because none of the arguments of a transitive verb belongs morphosyntactically nearer to the verb than the other, even if one can rightly state that these arguments are ordered semantically (and thus can use the notions of logical subject and object). Bailin & Grafstein (1991) and Dahlstrom (1991:99) showed that the Algonquian languages also violate the *binding principle* which requires that an antecedent must either c-command or linearly precede an anaphor in order to bind it (see Wunderlich 2005).

3. Morpho-syntax vs. dislocation syntax. Observations regarding morphosyntactic diversity make it doubtful whether morphology and syntax emerged at the same time. As is well-known, morphology and syntax can in principle do the same job, so that one of them seems superfluous. Indeed some languages nearly lack morphology except compounding (isolating languages such as Vietnamese), while others nearly lack syntax except juxtaposition (polysynthetic languages such as Yimas of New Guinea, Nivkh of Siberia, Greenlandic etc). These two types of languages, isolating vs. polysynthetic, are nearly incommensurable; no one can characterize the sort of language from which both could have developed.

Creole languages, created under the influence of poor input from a pidgin, are nearly isolating, a fact that prompted Bickerton (1981, and later) to assume that they show most clearly the influence of UG. Similarly, Klein & Perdue (1997) claim that UG is involved in the Basic Variety created by foreign

workers in Europe, which has a poor syntax and no morphology. These authors believe that linguistic compositionality started with simple syntax.

Polysynthesis on the other hand is centered around the concept of head-marking, including noun incorporation and verb serialization, and also involves bound morphemes with adverbial or attributive functions as well as various sorts of fusion. Polysynthesis itself seems to be a late product of repeated grammaticalizations (Mattissen 2004, 2006), but it deserves to be stated that only languages with head-marking are susceptible to such fusion. In general, head-marking languages have rich morphology but quite simple syntax. Since affixes on the verb represent the arguments, an inflected verb already represents a full clause. Independent nouns or NPs only serve for more explicit referential specification, and have the syntactic status of free adjuncts.

A comparison between morphology and syntax reveals that morphology is less flexible than syntax because it lacks dislocation, and thus can be regarded as poorer than syntax. (i) A morphological complex is characterized by fixed positions (except in some forms of polysynthesis), and no dislocation of elements, no agreement between elements, and no assignment of focus or topic takes place. (ii) A morphological complex is affected by more phonological rules, and thus susceptible to more irregularities, than a syntactic combination. (iii) A memorized morphological complex is more rapidly processed than a syntactic combination. (iv) Morphology is harder to acquire by adult learners than syntax. These observations make it highly improbable that syntax preceded morphology. On the contrary, if morphology tends to be more irregular, faster processed, and more difficult to be learned one rather expects morphology to be prior to syntax.

For thousands of generations, linguistic communities were quite small and isolated groups of hunter-gatherers; with a size between 150 and 1000 individuals, slightly increasing towards the end of the ice age. More people couldn't find their living in an area of several hundred square miles, unless the circumstances were extremely favourable. Seasonal trading between such groups probably existed but did not give the opportunity of much contact. Only at the end of the ice age 12,000 years ago did the living conditions highly improve, the size of communities increased, the gatherers settled, and finally agriculture was invented at various places around the world, which in turn caused a rapid increase of the population and forced new societal forms.

The distribution of languages resulting from this neolithic transition shows a mixture of small and large linguistic communities, and of elder and more recently developed linguistic profiles. Head-marking morphology could be a feature of phase 1 (before the neolithic transition), which survived in phase 2 (after the neolithic transition), whereas dislocation syntax is not necessarily a feature of phase 1 but could belong to phase 2. Evidence comes from several sources.

Nichols (1992), distinguishing between spread zones (where languages or language families rapidly spread, serving as lingua franca, and consequent languages succeeded, and little genetic and low structural diversity is found) and residual zones (often at the periphery of spread zones, with high genetic and high structural diversity, and no lingua franca besides local bilingualism), found high correlations between features of head-marking and residual zones. Similarly, Dahl's (2006) statistical survey shows that languages of the hunter zone ("traditionally spoken in areas not fully affected by the neolithic transition") have more morphological complexities and less VO order than those of the farmer zone ("traditionally spoken in areas with fully established agriculture"), and that the former do not include isolating languages. Dahl also observed that the speaker density (speakers per sq. km) is much lower in the New World (≤ 1) than in the Old World (20-150). A density of 1 speaker/sq.km means that a group of 1,000 people occupies a region of 32 kms at each side, so that external contacts must be rare.

Trudgill (2004: 306) points out that "small, isolated, low-contact communities with tight social network structures" have large amounts of shared information in common, are more likely to demonstrate complexities and irregularities, and to ensure the transmission of linguistic complexity from one generation to another, i.e., to show a slower rate of linguistic change, whereas „communities involved in large amounts of language contact [...] are likely to demonstrate linguistic pidginisation, including simplification, as a result of imperfect language learning." Irregular and non-transparent forms cause particular problems of memory load for adult learners; their mode of imperfect learning leads to regularisation of irregularities, to an increase in transparency, and to an increase in analytic over synthetic structures.

For the present purpose it is essential to keep in mind that speakers of a small community have large amounts of shared information, including memorized linguistic forms, which can rapidly be processed in repeated as well as ritual encounters. Such a situation is changed if the community is growing, partitioned into various specialized or spatially separated groups, so that less information is shared, the social networks become more pervious, and the probability of external contacts increases. More variation of interaction settings forces the speech act participants to use more specifications by independent NPs, to be more explicit regarding topic and focus, to use more transparent combinations rather than memorized forms, and to use forms that are less prone to irregularities, so that the input for the respective next learner generation gets modified. In addition, a spreading population experiences more external language contacts. Adults using the dominating language as a lingua franca produce simplified and at the same time more transparent varieties, which accelerates the process of language change.

Competition between independent demands forces the emergence of positional variants (i.e. dislocation). For instance, the requirement of realizing grammatical functions positionally (subject precedes object) and the requirement of realizing discourse functions positionally (topic precedes focus) can conflict with each other, and therefore various orderings such as SVO, O_{top}SV, and $O_{top}VS_{foc}$ can emerge. Something of this might have happened to the previous languages with rich morphology. My hypothesis, then, is:

> Syntax (in the sense that it enables flexible ordering by dislocation of elements, and that it displays particular positions for topic and focus, and eventually dependent marking such as morphological case) is a product of cultural evolution (that is, of iterated learning under the influence of cultural factors). Syntax emerges in a linguistic community with high variation of interaction settings, which is more likely to take place in languages of the farmer zone than in those of the hunter zone.

Of course, not all linguistic communities were involved in the agricultural expansion, or in the rapid change towards more syntactic patterning, and the economic and the linguistic factors not always coincided. Indeed many small languages with rich morphology remained, so that the linguistic diversity of today partly originates from more recent changes, and partly preserves older stages. These different origins explain why such incommensurable language types such as the isolating and the polysynthetic ones exist side by side. The present coexistence of dislocating syntax and head-marking morphology thus opens the window to different time depths.

4. Syntactic universals. Do we have to assume specific universals in order to explain the emergence of dislocating syntax? Hurford's (2000) and Kirby's (2002) simulations with iterated learning models show that a stable compositional (recursive) syntax can arise within thousands of generations given completely unstructured strings of symbols at the beginning, a learning algorithm that can induce heuristically-driven grammars, and the ability of agents to associate strings with complex meanings. These simulations did not make any distinction between morphology and syntax. If speakers already used head-marking regularly, a syntax with the dislocation property, or even an isolating language, certainly could emerge from that state within fewer generations, given certain preconditions that make the starting state unstable (for instance, imperfect learning). However, I do not know of any simulation that aimed at transforming a given linguistic type into another.

A condition of these simulations was that speakers can assign complex meanings, which is realistic insofar as one assumes that already early homo was likely to entertain complex thoughts. If the notion of a mental attitude with a propositional object (such as 'believe') is available, recursion is available, too. Corresponding mental attitude verbs could have been present at the time 35,000 years ago when art appeared.

The heuristics of a grammar-inducing learning algorithm can be compared with processing principles or with constraints in other frameworks. Since the beginning of the Minimalist Program (Chomsky 1995) universal principles have been reduced, and nowadays, some of these principles have been argued to be correlated with or grammaticalized from processing principles (Fanselow, Kliegl & Schlesewsky 1999, Hawkins 2004), characterizing how the brain works if confronted with linguistic input. Constraints dealing with dislocation can be adapted from the geometric system serving for optimal perception of motion (figure-ground, locality, transformations, traces). Constraints have competing demands; and only

their respective ranking regulates the actual balance between maximal expressivity (explicitness, distinctiveness, transparency) and minimal expense (rapid processing, use of contextual cues).

Dislocation syntax, then, can have emerged by some re-ranking of constraints, for instance, that a specified NP became better than a pronominal affix, and that the realization of topic and focus became high-ranked. This re-ranking was forced by external factors. The invention of dislocation was an advantage because more forms could be taken into consideration; it thus expanded the grammatical space. In the end, new constructions took over the task of elder ones. In other words, syntax forces less-flexible and redundant sorts of morphology to disappear.

References

Arbib, Michael. 2005. The mirror system hypothesis: How did protolanguage evolve? In Tallerman, 21-47.

Bailin, Alan & Ann Grafstein. 1991. The assignment of thematic roles in Ojibwa. *Linguistics* 29, 397-422.

Bickerton, Derek. 1981. *Roots of language*. Ann Arbor, Mich.: Karoma.

Briscoe, Ted. 2003. Grammatical assimilation. In Christensen & Kirby, 295-316.

Calvin, William H. 1991. *The ascent of mind. Ice age climates and the evolution of intelligence.* New York: Bantam Books.

Carstairs-McCarthy, Andrew. 2005. The evolutionary origin of morphology. In Tallerman, 166-184.

Chomsky, Noam. 1995. *The minimalist program*. MIT Press.

Chomsky, Noam. 2005. On phases. Ms.

Christensen, Morton H. & Simon Kirby (eds.) 2003. *Language evolution*. Oxford UP.

Corballis, Michael C. 2003. From hand to mouth: The gestural origins of language. In Christensen & Kirby, 201-218.

Dahl, Östen. 2006. In search of typological diversity. Symposium *Current topics in typology*, FU Berlin 14.1.06.

Dahlstrom, Amy. 1991. *Plains Cree Morphosyntax*. New York: Garland.

Enard, Wolfgang et al. 2002a. Intra- and interspecific variation in primate gene expression patterns. *Science* 296: 340-343.

Enard, Wolfgang et al. 2002b. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418: 869-872.

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford University Press.

Hockett, Charles. 1960. The origin of speech. *Scientific American* 203: 88-96.

Hockett, Charles. 1966. The problem of universals in language. In Greenberg (ed.) *Universals of language*, 1-29. MIT Press.

Hurford, James R. 2000. The emergence of syntax. In C. Knight, M. Studdert-Kennedy, and J.R. Hurford (eds.) *The evolutionary emergence of language: Social function and the origins of linguistic form*, 219-230. Cambridge UP

Kirby, Simon. 2002. Learning, bottlenecks, and the evolution of recursive syntax. In E. Briscoe (ed.) *Linguistic evolution through language acquisition: Formal and computational models*, 173-204. Cambridge UP.

Klein, Wolfgang & Clive Perdue. 1997. The Basic Variety, or: Couldn't language be much simpler? *Second Language Research* 13: 301-347.

Lai, Cecilia S. et al. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413: 519-523.

Mattissen, Johanna. 2004. A structural typology of polysynthesis. *Word* 55: 189–216.

Mattissen, Johanna. 2006. The ontology and diachrony of polysynthesis. In *Advances in the theory of the lexicon*, D. Wunderlich (ed.). Mouton de Gruyter, to appear.

Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago UP.

Oudeyer, Pierre-Yves. 2005. From holistic to discrete speech sounds: the blind snowflake-maker hypothesis. In Tallerman, 68-99.

Rizzolatti, Giacomo, Luciani Fadiga, Vittorio Gallese, & Leonardo Fogassi. 1996. Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* 3: 131-141.

Rizzolatti, Giacomo & M. A. Arbib. 1998. Language within our grasp. *Trends in Neuroscience* 21: 188-194.

Studdert-Kennedy, Michael. 2005. How did language go discrete? In Tallerman, 48-67.

Tallerman, Maggie (ed.) 2005. *Language origins. Perspectives on evolution*. Oxford UP.

Trudgill, Peter. 2004. Linguistic and social typology. *Linguistic Typology* 8: 305-320.

Wunderlich, Dieter. 2004. Why assume UG? *Studies in Language* 28: 615-641.

Wunderlich, Dieter. 2005. The challenge by inverse morphology. *Lingue e Linguaggio* 4: 195-214.