# Language Type Frequency and Learnability. A Connectionist Appraisal.

**Ezra Van Everbroeck** (ezra@ucsd.edu)
UCSD Department of Linguistics; 9500 Gilman Drive
La Jolla, CA 92093 USA

**Abstract**

In this paper, I present experimental data bearing on the controversial issue of the possible relationship between the frequency of language types and how easily they can be learnt. Using simple, artificial languages which only differ with respect to the properties we are interested in, I show that there does appear to be a relationship of some kind, although not as strong as one might have hoped. In particular, if a language type can be learnt relatively easily, then the models fail to predict its actual frequency in the real world. On the other hand, the connectionist models provide evidence that the language types which are unattested or highly infrequent are also impossible or hard to learn.

## Introduction

It has been known for a long time that languages differ in the ways in which they express 'who did what to whom?'. The three most important dimensions of this variability are: first, the word order of Subject, Object and Verb; second, the presence or absence of markers on the Verb; and, third, the presence or absence of markers on the Subject and/or Object (Nichols, 1986). The sentences below illustrate how the three strategies work. All (a) examples mean the same thing, but they get the message across in a different way — notice that the word order of S, O and V in (2) and (3) remains constant from the (a) to the (b) sentence.

1. (a) The matador killed the bulls.
   (b) The bulls killed the matador.

2. (a) The matador-he the bulls-them killed.
   (b) The matador-him the bulls-they killed.

3. (a) Killed-he-them the matador the bulls.
   (b) Killed-they-him the matador the bulls.

The three dimensions are also theoretically independent from each other, so we can imagine languages which (redundantly) combine a fixed word order with verbal and nominal marking, languages with none of the three, as well as the six other logical possibilities.

Things get even more complex, however, because the three dimensions are not necessarily binary. For example, in word order alone there are already six possible orders of S, O and V, next to a 'free' word order type in which many different combinations of the words are usually possible, albeit normally with pragmatically distinct meanings (Payne, 1992). And when case markers are used on the Subject and Object, we can at least distinguish between accusative strategies, in which (nominative) S is always marked differently from (accusative) O, ergative strategies, in which the (absolutive) S of the intransitive clause is marked similarly to the (absolutive) O of the transitive clause but the transitive S has a different (ergative) marker (Van Valin, 1992), and unmarked (or null) strategies.

So, we actually have at least 42 different language types (i.e. 7 word orders * 3 types of nominal marking * 2 types of verbal marking). As careful study of this many real languages would threaten to take up a lifetime of research, the route taken here has been to create artificial languages instead, and to use connectionist models to investigate the latter in a systematic manner.

Given the plethora of possible language types, it is not surprising that they are not found with equal frequency in the world. For example, the types with a fixed word order of SOV account for about half of the world's languages, whereas the types with OSV may even be unattested (see also below). In general, there are also no known languages which consistently fall on the extremes of the three dimensions: i.e. which either don't have any such mechanism for signaling 'who did what to whom', or which simultaneously employ all three mechanisms in a single sentence. While there are good common-sense reasons for these last two facts — i.e. there has to be some strategy or all communication would fail, on the one hand, and useless redundancy is not likely when it wastes resources, on the other hand — the connectionist simulations to be presented below are an attempt to explain the data by bringing learnability issues into the picture (cf. Christiansen & Devlin 1997). The important questions are:

- Can a neural network learn the attested language types?
- Will a neural network fail to learn the unattested language types?
- Will a neural network learn the more frequent language types faster/better?

If the answer to all three were to be 'yes', we could claim that there is a strong causal relationship between frequency and learnability of language types. However, the experimental evidence presented here only warrants a weaker conclusion.

The structure of this paper is as follows: in the next section, I will present the available linguistic frequency data in more detail, so that the phenomena to be accounted for are clear. In section 3, I briefly go over the setup of the simulations. Section 4 presents the results of the various simulations. The last section, then, wraps up the paper and provides

some language acquisition evidence supporting the posited connection between learnability and frequency.

## Linguistic Frequency Data

Historically, the focus in language typological research has been on finding correlations between the word orders of various pairs like adposition — NP, genitive — noun, or NP — relative clause (see e.g. Greenberg 1963; Hawkins 1988; Dryer 1992). However, such correlations do not play a role in the simulations reported in this paper (but see Christiansen & Devlin (1997) and Van Everbroeck *(in prep)* for related work in which they do), because the sentences used only contain a Subject, Verb, and possibly an Object.

The information summarized in Table 1 below is much more relevant, then, as it shows the frequencies of the 6 possible fixed word orders of S, O and V (Tomlin, 1986; Dryer, 1989). One should keep in mind that these numbers control for historical and geographical biases, so that only unrelated languages are taken into consideration.[1]

Table 1: Language type frequencies.

| SOV | SVO | VSO | VOS | OVS | OSV |
|---|---|---|---|---|---|
| 51% | 23% | 10% | 9% | .75% | .25% |

It is not hard to see that SOV is by far the most frequent word order, with OVS and OSV being extremely rare — the latter may even be absent completely (Polinskaja, 1989). The percentages in Table 1 also show that Subject-before-Object languages are much more common than their Object-before-Subject counterparts — compare Greenberg's (1963: 77) Universal 1: "In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object." Moreover, the Subject-initial SOV and SVO are noticeably more frequent than either the Verb-initial or the Object-initial language types. Finally, the missing 6% in Table 1 accounts for free word order languages like Native American Klamath (Barker, 1964).

With regards to the other two dimensions, nominal marking and verbal marking, only limited frequency information is available (see Nichols 1986 for the best summary to date). In general, language types with redundant nominal and verbal marking, and language types with neither kind of marking (as in English), are less common than the other two possibilities. It also appears that most languages use some form of verbal marking, though usually with supplementary information being provided by either word order or nominal marking. With respect to the latter, it seems safe to say that accusative systems outnumber ergative systems by a considerable margin, though the presence of mixed systems (e.g. accusative for pronouns, but ergative for other nominals — see Morris 1998) again complicates matters considerably.

In summary, then, at least some of the frequency patterns which one would expect to find mirrored in the learnability findings below — if there is indeed a connection between frequency and learnability — are at various levels of abstractness. First, any of the three strategies should be sufficient in at least some cases. Second, the language types which do not make use of word order or any kind of marking should be unlearnable. Third, the ones that are highly redundant should be learnable, but should not offer much of an improvement over their less verbose learnable counterparts — there is a production and processing cost to verbosity (Kirby, 1997). Fourth, Subject-initial languages should be more easily learnable than Object-initial languages. Fifth, accusative language types should present fewer problems for the neural networks than ergative language types.

## Experimental Setup

The network model used to test the hypotheses just mentioned follows the following steps:

- Generate 42 artificial languages which only differ with respect to the three dimensions;
- Train an identical network on a corpus of sentences of each language;
- Compare 1) how well each language type is mastered, and 2) how well the network can generalize each time.

The artificial languages have been generated using simple context-free grammars which produce sentences appropriate to each language type. For example, the doubly redundant accusative SOV/VN grammar generates transitive SOV or intransitive SV sentences in which there are also verbal markers (indicated by the /V) as well as nominal markers (/N). (If a strategy is not used in a language type, X's are used instead of the other letters. So, VOS/XN is a VOS language with only nominal marking; XXX/VX is the free word order language type with only verbal marking. These mnemonic references will be used constantly below.)

The neural network architecture used in these simulations is illustrated in Figure 1. It is basically a simple recurrent neural network (Elman, 1992) with an extra recurrent layer at the output to provide it with more memory capacity. The

[1] We are not interested in the raw number of individual languages which exhibit a certain word order, partly because there is still no clear definition of what makes a language as opposed to a dialect, and partly because we do not want closely related languages to count individually: e.g. German and Dutch might as well have ended up as a single language if history had taken a different turn; similarly, due to their geographical proximity to one another, all the languages spoken in the Balkan have some linguistic features in common though they belong historically to different groups. If we only look at a single representative language from each larger family instead of counting individual languages, we have a much better chance of capturing a universal phenomenon which is independent of where a language is spoken.

task of this network, which sees one word per time step at the input layer, is to construct a representation at the output layer which shows which words it parses to be the Subject, Object or Verb.

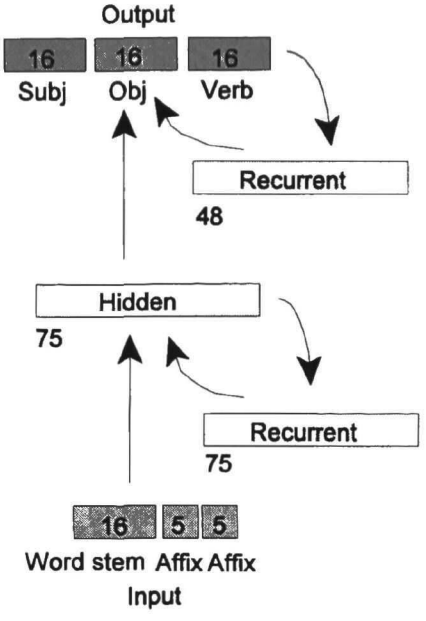


Figure 1: The network used in the simulations.

At the input layer, the 26 units encode a single lexical item (16 units with always exactly 6 units asserted) and up to two verbal or nominal suffixes (5 units each; 3 units asserted for a suffix). The corpus for each language contains 600 nouns and 100 verbs, with half of each category being used in the training set and the other half in the test set; the suffixes remain constant across the two sets.

At the output layer, there are three slots with 16 units — one slot each for S, O and V. Hence, as the network sees each of the words in a sentence similar to English 'They see me', it has to put the lexical item representation for 'they' in the first slot, then add that of 'see' to the third slot, and finally that of 'me' in the second. Each sentence is followed by a time step in which the pattern for the entire sentence has to be maintained, and then a reset signal is sent to the network. It is at the maintenance time step that the performance of the network on each sentence is calculated. For a sentence to be correct, the pattern of activation in each of the three output slots has to be closest in Euclidean space to the target word. As soon as a single word does not match the correct output, the entire sentence is considered incorrect.

For training, each network sees a corpus of 3,000 sentences of the relevant language (e.g. SVO/XN) for 10 epochs. The backpropagation algorithm is used to adjust the weights, with a learning rate of 0.15. For generalization testing, then, another corpus of 3,000 sentences (but with completely different nouns and verbs) is presented to the network and the performance determined. In the next section, the numbers reflect the percentages of sentences correctly processed on the training and the test set.

## Results and Discussion

### Fixed Word Order and Accusative Marking

The results of the simulations with the fixed word order language types with accusative N-marking are shown in Table 2 below. If we just look at the Subject-before-Object languages, we find excellent performance on the training set (>97%) as well as the test set (>93%, with one exception).

The 23.7% test set performance for SOV/XX, however, is actually a positive result, because it meshes well with Greenberg's (1963: 96) Language Universal 41, which specifies: "If in a language the verb follows both the nominal subject and the nominal object as the dominant order, the language almost always has a case system". Hence, SOV/XX is a language type which is extremely rare among the languages of the world, and the fact that the network has problems with the test set is compatible with the posited relationship between frequency and learnability.

Table 2: Percentages of sentences correctly analyzed (accusative N-marking; fixed word order languages).

| | | Train | Test | | Train | Test |
|---|---|---|---|---|---|---|
| VSO | XX | 100 | 99.0 | VOS | 75.9 | 28.0 |
| | XN | 100 | 99.6 | | 99.9 | 98.0 |
| | VX | 100 | 98.1 | | 80.0 | 50.0 |
| | VN | 100 | 98.7 | | 100 | 98.3 |
| SVO | XX | 99.9 | 95.5 | OVS | 81.2 | 12.4 |
| | XN | 99.9 | 95.9 | | 93.3 | 87.5 |
| | VX | 99.9 | 97.0 | | 96.9 | 91.9 |
| | VN | 99.9 | 97.9 | | 99.9 | 99.2 |
| SOV | XX | 84.4 | 23.7 | OSV | 24.9 | 16.3 |
| | XN | 98.4 | 97.3 | | 99.9 | 99.1 |
| | VX | 97.4 | 93.9 | | 27.2 | 17.1 |
| | VN | 100 | 98.8 | | 99.0 | 99.0 |

It turns out that there is a very good reason for the bad performance on the SOV/XX test set: when one 'hears' a sentence with all new words, and the words are also unmarked (i.e. there is no marker distinguishing even the nouns from the verbs), it is impossible to know whether the second word in the sentence is the Object of a transitive clause (in SOV), or the Verb of an intransitive clause (in SV) — all one can know is that it is the second word in the

sentence. Hence, the network hedges its bets and spreads partial activation over both the Object and the Verb units at the output. Though in theory the presence/absence of a third word in the sentence can disambiguate between the two options — if another word follows, it must be the Verb, so the second word must have been the Object — the network fails to recover sufficiently (i.e. remove activation from the incorrect units and fully activate the correct ones) by the time performance is determined.[2]

If we now turn our attention to the twelve Object-before-Subject language types, we find that the networks in general have a harder time learning them — especially generalization is often problematic. As with SOV/XX, the language types without any kind of marking are very bad performers, and they are also infrequent or unattested. If verbal marking is added, only OVS benefits enough to produce a useful communicative system (>90% on training and test set), because the Verb now separates the two ambiguous nominals. With nominal case-marking, on the other hand, all the resulting systems appear learnable, be they /XN or /VN. Still, as with the Subject-before-Object language types, we again find that the /VN types tend to have little or no advantage over their /XN counterparts — a finding which is expected if we assume that speakers/hearers would prefer to avoid overly redundant and verbose systems.

It is also worth pointing out that OSV/XX and OSV/VX are the only two language types for which not even the training set is learnt successfully. OSV/VN is unrealistic in that is highly redundant. This leaves us with OSV/XN as an apparently viable type for the network, though it may be unattested among natural languages. A possible explanation here may be that case systems tend to degrade over time (Venneman, 1975), which would result in unlearnable languages with OSV. Hence, this basic word order would automatically become extinct over time.

Finally, I want to return to the general frequency hierarchy given earlier in Table 1, and its counterpart in Table 2. It turns out that one can find interesting parallels between this hierarchy on the one hand, and a comparison of the absolute order of Subject, Object and Verb in transitive and intransitive sentences, on the other hand. This is diagramed in Table 3 below.

Table 3: Relationship between transitive and intransitive clauses in terms of absolute word order.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| VSO | V | S | O | VOS | V | O | S |
| | V | S | | | V | S | |
| | ✓ | ✓ | | | ✓ | ✗ | |
| SVO | S | V | O | OVS | O | V | S |
| | S | V | | | V | S | |
| | ✓ | ✓ | | | ✗ | ✗ | |
| SOV | S | O | V | OSV | O | S | V |
| | S | V | | | S | V | |
| | ✓ | ✗ | | | ✗ | ✗ | |

Notice that with VSO and SVO, the fixed word order is actually most useful in that it guarantees that the Verb and Subject will always appear in the same slot — thereby facilitating processing. With SOV and VOS, only the first word remains constant; with the infrequent OVS and OSV types, there is no overlap between transitive and intransitive sentences at all. Table 3, however, does not explain at all why SOV actually accounts for 51% of the world's language types, and VSO for only 10%. I will return to this quandary in the general discussion, but let me simply mention that the psychological primacy of agentive/animate Subjects is likely involved.

## Fixed Word Order and Ergative Marking

Let us now take a look at the results for the ergative language types in Table 4. Recall that in an ergative language, the O of the transitive clauses is treated similarly to the S of the intransitive clause (both are semantical Patients most of the time), whereas the S of the transitive clause (usually the Agent) receives different marking. For the Subject-before-Object language types, we find an almost identical picture to the one in Table 2 above. All language types are learnable and can be generalized from quite well, except for the rare language type SOV/XX. (Obviously, the /XX language types are really identical to the ones from the accusative set. They are repeated here for completeness, and to give some idea of the range of variation between different simulations of the same type.)

The fact that the Subject-before-Object ergative language types can be learnt easily is an encouraging finding, because such languages do certainly exist. When we turn to the Object-before-Subject languages, however, we find a different picture — even a cursory glance suffices to see that the percentages are generally much lower than those for the accusative counterparts. But for three OVS types, performance on the training set is down, and generalization also turns out to be quite problematic in almost all cases. For OVS, the redeeming feature appears to be that the Verb separates the nouns, which makes it easier to tell S, O and V apart. With the VOS types, the networks are rote learning

[2] This finding demonstrates that the simple recurrent network architecture used here does not perform well when there is a heavy memory load — i.e. it has to store the exact identity of the second word in the recurrent layer when it waits for the third word. An architecture in which such a word could be stored in a dedicated buffer would obviously have a better chance of dealing with this task. Alternatively, the types SOV/VX and SOV/XN solve the problem by disambiguating the word forms through the use of morphological markers. One should also keep in mind that the generalization task which the network faces is unrealistically hard in that it has to process sentences in which *all* the words are completely unknown.

the training set; a strategy which fails miserably when the new words in the test set are presented. And things are even worse for the — luckily unattested — OSV language types, because these networks fail to even pick up anything useful in the training set.

Table 4: Percentages of sentences correctly analyzed (ergative N-marking; fixed word order languages).

| | | Train | Test | | Train | Test |
|---|---|---|---|---|---|---|
| VSO | XX | 99.9 | 98.0 | VOS | 84.8 | 33.1 |
| | XN | 99.9 | 98.6 | | 66.9 | 40.5 |
| | VX | 99.9 | 98.6 | | 76.9 | 42.0 |
| | VN | 99.9 | 98.6 | | 79.2 | 49.5 |
| SVO | XX | 99.9 | 96.8 | OVS | 73.2 | 14.4 |
| | XN | 99.9 | 96.1 | | 98.5 | 96.6 |
| | VX | 100 | 96.7 | | 97.9 | 92.3 |
| | VN | 99.9 | 97.5 | | 99.8 | 99.0 |
| SOV | XX | 80.5 | 21.7 | OSV | 12.9 | 4.2 |
| | XN | 98.3 | 98.3 | | 12.5 | 4.4 |
| | VX | 98.2 | 95.4 | | 11.8 | 4.1 |
| | VN | 99.9 | 99.4 | | 12.2 | 3.4 |

What is so difficult about ergative nominal marking and Object-before-Subject languages? In short, the case markers are no longer helpful for telling the Subject and Object apart. With an accusative system, an accusative always signals an Object; the nominative always a Subject. But in an ergative system, Subjects can be marked either with the absolutive (intransitive) or the ergative (transitive) marker. Objects will always appear with the absolutive case, but the ambiguity has already been introduced into the system. And as before, the networks fail to recover from such an initially ambiguous input sequence: in theory, the presence or absence of an ergative form should allow the network to figure out the grammatical role of the absolutive marked noun, but this does not seem to happen.[3]

## Free Word Order

The final set of results are shown in Table 5. They are for the free word order languages, both accusative and ergative.

[3] Additional training with the current network does not alleviate this problem, but Bill Morris (personal communication) has found that a much larger Elman net can be taught to handle such patterns much better. Still, one would expect language types which require more resources than others to be at a competitive disadvantage when children have to acquire them, and, therefore, to be either absent, or very rare — cf. Kirby 1997.

Table 5: Percentages of sentences correctly analyzed (accusative/ergative marking; free word order languages).

| Acc | | Train | Test | Erg | Train | Test |
|---|---|---|---|---|---|---|
| XXX | XX | 15.0 | 2.3 | XXX | 11.5 | 1.5 |
| | XN | 84.7 | 75.1 | | 32.2 | 25.8 |
| | VX | 45.3 | 33.8 | | 40.5 | 29.1 |
| | VN | 99.0 | 94.0 | | 38.3 | 19.2 |

It is easy to see that free word order languages as implemented here — i.e. with a completely random word order — are generally harder to learn than their fixed word order counterparts. Only the accusative XXX/VN type performs adequately. However, ergative XXX/VN is actually also attested, though the simulations would not predict this. Although the answer may lie in the overly random nature of the free word order simulated here, this issue does require further attention.

## General Discussion and Conclusion

The motivation of this paper has been to explore the possible relationship between frequency of language types, and their learnability. We have seen that four of the five predictions formulated earlier find support in the connectionist simulations: first, the language types which do not use any of the three (i.e. XXX/XX) are unlearnable. Second, language types which are highly redundant (e.g. SVO/VN) do not have a significant advantage — if at all — over the simpler SVO/VX or SVO/XN, so the latter should be preferred by processing cost sensitive mechanisms. Third, the Subject-before-Object language types are indeed much easier to learn than their less frequent Object-before-Subject counterparts. Fourth, accusative language types as a group are apparently easier for a simple recurrent neural network than similar ergative types. With respect to the fifth prediction, namely that all three strategies (word order, verb marking, nominal marking) should be independently capable of expressing 'who did what to whom', we have found that this bit of conventional linguistic wisdom is not supported by the simulations: XXX/VX is not a learnable type. On the other hand, adding verbal marking to a fixed word order or nominal marking does make the task of the network noticeably easier in a number of the simulations.

Other findings which are compatible with the typological data are the bad performance on SOV/XX (Greenberg, 1963); the fact that V-initial languages prefer verbal marking (Nichols, 1986); and the fact that free word order languages generally occur with case systems (Payne, 1992).

However, we have also seen that the network results do not map perfectly onto the frequencies observed in the real world: e.g. the VSO networks generally perform much better than their frequency would lead one to expect. Similarly, some of the Object-before-Subject language

types, or the ergative types are not as frequent as the modeling results predict. Also, there is at least one language type, ergative XXX/VN, which is attested but on which the model performs badly.

There are two different approaches to these criticisms: one is to make the models more complex — e.g. by adding semantics and perceptual salience for Subjects (Langacker, 1993), one could build in a psychologically plausible bias for SVO and SOV. The other one is to take the network results at face value and use them as a heuristic for further investigation: if a language type appears to be unlearnable, then it may employ other, compensatory mechanisms (e.g. tone) to express 'who did what to whom'. The simulations tell us where to look for such mechanisms.

So, how tight is the relationship between frequency and learnability? The simulations presented here do not justify a strong causal connection between the two, but they do lend support to a weaker position. Namely, they can tell us which language types are unattested because they are impossible to learn and which ones are unattested because they are overly redundant. If a language type is attested, however, then the networks fail to predict its frequency. It seems safe to assume that other factors, more historical and geographical in nature, determine the actual frequency of a language type, just as they determine the number of the speakers of any given language.

This conclusion should not really come as a surprise. Christiansen & Devlin (1997) found that their neural networks had a harder time learning language types with head-dependent word order inconsistencies — exactly the types which are also less frequent in the real world. And some time ago, Slobin & Bever (1982) reported that children acquiring different language types do indeed differ significantly in how fast they learn to understand 'who did what to whom' in their native languages. For example, infants learning Turkish (a language with abundant and unambiguous nominal marking) are much better at this task than infants acquiring Serbo-Croatian (which requires attention to word order and often ambiguous case markers). Children are somewhat like neural networks then, at least with respect to this task.

## Acknowledgments

I would like to thank Chris Barker, Gary Cottrell, Ron Langacker, Bill Morris and Masha Polinsky, as well as an anonymous reviewer for their comments on an earlier version of this paper. Numerous useful suggestions were also made by the audiences who sat through talks at the Center for Research in Language and the AI Research Group. The early stages of the research presented here were performed when the author was a fellow of the Belgian American Educational Foundation.

## References

Barker, M. (1964), *Klamath Grammar*, Berkeley: UC Press

Christiansen, M. and J. Devlin (1997), Recursive Inconsistencies Are Hard to Learn: A Connectionist Perspective on Universal Word Order Correlations, in *Proceedings of the 19th Annual Cognitive Science Society Conference*, Mahwah, NJ: Lawrence Erlbaum, 113-118

Dryer, M. (1989), Large Linguistic Areas and Language Sampling, in *Studies in Language* 13.2, 257-292

Dryer, M. (1992), The Greenbergian Word Order Correlations, in *Language* 68.1, 81-138

Elman, J.L. (1992), Grammatical Structure and Distributed Representations, in S. Davis (Ed.) *Connectionism: Theory and Practice*. New York: Oxford University Press

Greenberg, J. (1963), Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements, in J. Greenberg (Ed.) *Universals of Language*. Cambridge, MA: MIT Press

Hawkins, J. (1988), Explaining Language Universals, in J. Hawkins (Ed.) *Explaining Language Universals*. Oxford: Basil Blackwell

Kirby, S. (1997), Competing motivations and emergence: explaining implicational hierarchies, in *Language Typology* 1, 5-32

Langacker, R. (1993), Reference-point constructions, in *Cognitive Linguistics* 4.1, 1-38

Morris, W. (1998), *Emergent Grammatical Relations: An Inductive Learning System*, unpublished doctoral dissertation, UC San Diego

Nichols, J. (1986), Head-marking and dependent-marking grammar, *Language* 62.1, 56-119

Payne, D. (1992), Introduction, in D. Payne (Ed.) *Pragmatics of Word Order Flexibility*. Amsterdam: John Benjamins

Polinskaja, M. (1989), Object initiality: OSV, in *Linguistics* 27, 257-303

Slobin, D. and T. Bever (1982), Children use canonical sentence schemas: A crosslinguistic study of word order and inflections, in *Cognition* 12, 229-265

Tomlin, R. (1986), *Basic Word Order. Functional Principles*, London: Croom Helm

Van Everbroeck (in prep.), Syntax vs. Morphology

Van Valin,R. (1992), An Overview of Ergative Phenomena and Their Implications for Language Acquisition, in D. Slobin (Ed.) *The Crosslinguistic Study of Language Acquisition. Volume 3*. Hillsdale, NJ: Lawrence Erlbaum

Vennemann, T. (1975), An Explanation of Drift, in C. Li (Ed.) *Word Order and Word Order Change*. Austin, TX: University of Texas Press