A Connectionist Simulation of the Empirical Acquisition of Grammatical Relations

William C. Morris[1], Garrison W. Cottrell[1], and Jeffrey Elman[2]

[1] Computer Science and Engineering Department
University of California, San Diego
9500 Gilman Dr., La Jolla CA 92093-0114 USA
[2] Center for Research in Language
Department of Cognitive Science
University of California, San Diego
9500 Gilman Dr., La Jolla CA 92093-0114 USA

Abstract. This paper proposes an account of the acquisition of grammatical relations using the basic concepts of connectionism and a construction-based theory of grammar. Many previous accounts of first-language acquisition assume that grammatical relations (e.g., the grammatical subject and object of a sentence) and linking rules are universal and innate; this is necessary to provide a first set of assumptions in the target language to allow deductive processes to test hypotheses and/or set parameters.

In contrast to this approach, we propose that grammatical relations emerge rather late in the language-learning process. Our theoretical proposal is based on two observations. First, early production of childhood speech is formulaic and becomes systematic in a progressive fashion. Second, grammatical relations themselves are family-resemblance categories that cannot be described by a single parameter. This leads to the notion that grammatical relations are learned in a bottom up fashion. Combining this theoretical position with the notion that the main purpose of language is communication, we demonstrate the emergence of the notion of "subject" in a simple recurrent network that learns to map from sentences to semantic roles. We analyze the hidden layer representations of the emergent subject, and demonstrate that these representations correspond to a radially–structured category. We also claim that the pattern of generalization and undergeneralization demonstrated by the network conforms to what we expect from the data on children's generalizations.

1 Introduction

Grammatical relations are frequently a problem for language acquisition systems. In one sense they represent the most abstract aspect of language; subjects transcend all semantic restrictions – virtually any semantic role can be a subject. While semantics is seen as being related to world-knowledge, syntax is seen as existing on a distinct plane. For this reason there are language theories in which grammatical relations are considered the most fundamental aspect of language. One approach to learning syntax has been to relegate grammatical relations and their behaviors to the "innate endowment" that each child is born with. There are a number of theories of language acquisition (e.g., [4, 27, 46, 47]) that start with the assumption that syntax is a separate component of language, and that the acquisition of syntax is largely independent of semantic considerations. Accordingly, in these theories there is an innate, skeletal syntactic system present from the very beginning of multiword speech. Acquiring syntax consists of modifying and elaborating the skeletal system to match the target language.

This assumption of innate syntax inevitably leads to a problem, sometimes referred to as the "bootstrapping problem". How does one *start* this "purely syntactic" analysis? How does one start making initial assignments of words to grammatical relations (i.e., subject, object, etc.)? A commonly proposed mechanism involves the child tentatively assigning nominals to grammatical relations based on their semantic content by linking rules[1] (e.g., Pinker [46, 47]). This implies that these grammatical relations and linking rules are present at the very beginning of the learning process.

One problem with this approach is that cross-linguistically the behaviors of grammatical relations differ too much to be accommodated by a single system. Proposals have been put forward [36, 47] that a single parameter with a binary value ("accusative" or "ergative") is sufficient to account for the extant grammatical systems. This has been shown to be inadequate [29, 40, 51] because there are languages that have neither strictly accusative nor strictly ergative syntax.

We propose a language acquisition system that does not rely on innate linguistic knowledge [40]. The proposal is based on Construction Grammar [24, 25] and on the learning mechanisms of PDP-style connectionism [50]. We have hypothesized that abstractions such as "subject" emerge through rote learning of particular constructions, followed by the merging of these "mini-grammars". The claim is that in using this sort of a language acquisition system it is possible for a child to learn grammatical relations over time, and in the process accommodate to whatever language-specific behaviors his target language exhibits.

Here we present a preliminary study showing that a neural net that is trained with the task of assigning semantic roles to sentence constituents can acquire grammatical relations. We have demonstrated this in two ways: by showing that this network associates particular subjecthood properties with the appropriate verb arguments, and by showing that the network has gone some distance toward abstracting this nominal away from its semantic content.

In the following, we first review the ways in which the grammatical relation "subject" appears in several languages. This gives rise to the notion that grammatical relations do not have, for example, only two patterns of ways in which they control (in the linguistic sense) other categories. Rather, grammatical relations exhibit a variety of patterns of control over syntactic properties. This suggests it would be difficult for the subject relation to be described by a binary innate parameter. Next, we review relevant developmental data on the acquisition of syntax. The evidence we review suggests that 1) syntax is acquired in a bottom up, data-driven fashion, and 2) that there are specific

---

[1] Linking rules are heuristics (or algorithms, depending on the theory) for making provisional assignments of verb arguments to grammatical relations. The criteria for the assignments are semantic. Because virtually any semantic role can be a subject, the algorithmic variants of these theories are quite complicated. For a recent treatment of linking rules, see Dowty [21].

patterns of over- and under- generalization that reflect the nature of the linguistic input to the child. We then review the theory proposed by Morris [40] based on this data. Finally, we present a connectionist simulation of one stage of the theory, and demonstrate that the system acquires a notion of "subject" without any innate bias to do so.

2 The Shape of Grammatical Relations

While a number of theorists have explored the real complexity of grammatical relations (e.g., [19, 20, 23, 29, 51]), there remains a perception among some theorists (e.g., [34, 35, 36]) that grammatical relations are essentially a binary phenomenon: grammatical relations are deemed to be either accusative or ergative, and hence an "ergative parameter" determines the behaviors. This has been the prevailing view in a number of language acquisition theories [47].

A first-order approximation of the difference between accusative and ergative grammatical relations is that the subject of a syntactically accusative language is typically the agent of an action, while in a syntactically ergative language the "subject",[2] or subject-like grammatical relation, is typically the patient of an action. One potential distinguishing property (indicative, though not decisive) would be which nominal in a sentence controls clause coordination. Thus in the sentence, *Max hit Larry and ran away,* who ran away? In a strongly syntactically accusative language, it is Max that ran away; in a strongly syntactically ergative language, it is Larry that ran away.

For those who regard the accusative/ergative split as being simply binary, the problem becomes merely identifying the subject. If the subject is the agent, then the language is accusative, if it is the patient, it is ergative. But the problem is not that simple. It is not merely the identity of the subject that is the issue, but what properties do the various grammatical relations control? In some sense, the question is what "shape" do the grammatical relations in a language take on?

We have examined the literature to find the syntactic properties that are associated with subjects cross-linguistically. Perhaps the definitive work in this area is Keenan [29], from which we have extracted a set of six properties that are capable of being associated with subjects (and quasi-subjects) cross-linguistically:

1. Addressee of imperatives.
2. Control of reflexivization. E.g., *Max shaved himself.* (The controller of the reflexive is the subject.)
3. Control of coordination. E.g., *Max pinched Lola and fled.* (The deleted argument of the second clause is coreferential with the subject of the first clause.)
4. Target of equi-NP deletion. E.g., *Max convinced Lola to be examined by the doctor. Max convinced the doctor to examine Lola.* (The deleted argument of the embedded clause is the subject.)
5. Ability to launch floating quantifiers. E.g., *The boys could all hear the mosquitoes.* (The quantifier *all* refers to the subject, i.e., *boys*, rather than to the object, i.e., *mosquitoes.*)

---

[2] Because of the associations with accusative phenomena carried by the term "subject" in a number of theoretical approaches, one might wish to call the primary grammatical relation in syntactically ergative languages something else. The term "pivot" has been used.

6. Target of relativization deletion. E.g., *I know the man who saw Max. I know the man who Max saw.*

In English the last item is a free property; *any* nominal that is coreferential with the relativized matrix nominal can be deleted in the embedded clause in relativization. The examples demonstrate two of the cases.

The grammatical relations of various languages control various combinations of these (and other) properties. This is what we mean by the "shape" of grammatical relations. We have analyzed these syntactic properties in English and in two other languages, Dyirbal (Australian) [18, 20] and Kapampangan (Philippine), which have rather different constellations of properties from those of English, as well as from each other [40]. Grammatical relations in these languages have shown interesting patterns of behavior. For example, in English the first five of these properties are controlled by the subject, the last is a "free property", not controlled by any grammatical relation. In Dyirbal, properties 3, 4, & 6 are controlled by an "ergative subject", or "pivot" [18, 20]. In Kapampangan, one grammatical relation (which tends to be the agent) controls properties 1, 2, & 3, while another (which ranges over all semantic roles) controls properties 5 & 6. Property 4 can be controlled by either of the grammatical relations.

Hence English is a highly syntactically-accusative language, Dyirbal is a highly syntactically-ergative language, and Kapampangan appears to be a split language, neither highly ergative nor highly accusative in syntax. This is discussed at some length in [40], but as these languages do not bear directly on the present simulation, we will simply note that this issue is addressed in both the theoretical proposal and in our long-term goals.

Our purpose for raising the issue here is to argue that for a language acquisition to be "universal", i.e., capable of learning any human language, it must be able to accommodate a variety of language types. Simply settling on the identity of the subject is not sufficient. Rather, the various control patterns ("shapes") described above must be accommodated. Our proposal involves a system that can learn a variety of shapes.

3 Review of Data from Psycholinguistic Studies

There are several avenues of psycholinguistic data that we have explored. One of these is the issue of early abstraction vs. rote behavior. There have been a number of studies that have indicated that children's earliest multiword utterances have been largely rote or semi-rote behaviors [1, 2, 11, 12, 13, 44, 45, 53]. In a pair of studies Tomasello and Olguin showed an asymmetry between the relative facility with which two-year-old children can manipulate nouns, both in terms of morphology and syntax, and the relative difficulty with which they handle verbs. Tomasello & Olguin [55] demonstrated their productivity with nouns, while Olguin & Tomasello [43] showed their relative nonproductivity with verbs. It appears that the control that children have over verbs very early in the multi-word stage is largely rote; there is no systematic relationship between them. That is, there is little or no transfer from knowledge of one verb to the next.

There have been a number of studies [26, 41, 42, 56, 57] that have been interpreted as providing evidence of early abstraction. There are several problems with the interpretations of these studies, however. Some of these have interpreted arguably rote behaviors as representing abstraction [54], and others have interpreted small-scale systematic behavior as large scale systematic behavior [3, 15, 44, 45]). That is, it was found that certain systematic behaviors were limited to semantically similar predicates.

Despite the fact that an individual child's developing grammar is a quickly moving target, the issues of systematic and non-systematic behaviors can in certain instances be teased out. Indications of systematic behaviors can be seen in overgeneralization, and indications of the limits of systematic behaviors can be seen in undergeneralization.

In numerous studies, Bowerman [5, 6, 7, 8, 9, 10] has investigated instances of overgeneralization in child speech; overgeneralization is the phenomenon of extending rules inappropriately. For example, children exposed to English learn the "lexical causative" alternation, as in *the ball rolled∼Larry rolled the ball,* and *the vase broke∼Max broke the vase.* Children inappropriately extend this alternation to verbs such as *giggle* or *sweat* to produce such sentences as *Don't giggle me* or *It always sweats me* [9]. Overgeneralizations of this sort are evidence that the child has developed the notion of a class of verbs, such as *roll, float, break, sweat, giggle,* and *disappear,* which share a semantic role (patient) in their intransitive forms, and that the child is willing to treat them the same syntactically. The fact that this is inappropriate for the word *sweat* means that the child is extremely unlikely to have heard this usage before, therefore the child has used systematic behavior to produce this utterance. Another of Bowerman's studies [10] involved the overgeneralization of linking rules. Children rearranged verb-argument structures in accordance with a linking rule generalization rather than in accordance with some presumed verb-class alternation (e.g., *I saw a picture which enjoyed me.*).

Of particular note here is the timing of these, and other, overgeneralizations. Most of the overgeneralizations that Bowerman has studied, including the lexical causative overgeneralization discussed above, appear starting between two and a half and three and a half years of age. The linking rule overgeneralizations started appearing after the age of 6. The former overgeneralizations are presumably learned behaviors—the child must learn what sorts of verb classes exist in a language and what alternations are associated with them before these overgeneralizations can occur. On the other hand, according to many nativist theories, linking rules are innate [46, 47]. Furthermore, linking rules must be active very early in multi-word speech in order for the first tentative assignments of nouns to grammatical relations to be made, a necessary step in breaking into the syntactic system. Yet the overgeneralizations ascribable to linking rules do not appear until the age of six years or later.

If we can judge by overgeneralization, it would appear that linking rules are not innate; at the very least it appears that they are not active at a time when they are most needed, i.e., early in multi-word speech. The alternative is that they are not necessary precursors to multiword speech. Rather, they are highly abstract generalizations that first give evidence of existence after a large portion of the grammar of a language has been mastered.

Undergeneralization, too, has a role to play in determining the nature of the learning mechanisms. A number of studies have been conducted showing an interesting asymmetry in the learning of the passive construction in English. A study by Maratsos, Kuczaj, Fox, & Chalkley [38] showed that four- and five-year-old children could understand both the active and passive voices of action verbs (e.g., *drop, hold, shake, wash*), but had difficulty understanding the passive voices of psychological or perceptual verbs (e.g., *watch, know, like, remember*). Maratsos, Fox, Becker, & Chalkley [37] showed that this difficulty appeared to extend until the age of 10. Another studies by de Villiers et al. [17] confirmed the comprehension asymmetry between the two types of verbs, while a study by Pinker et al. [48] showed a similar asymmetry in production. In a preliminary study Maratsos et al. [37] also showed that parental input to children was limited in a similar way: parents used few, if any, experiential verbs in the passive voice. [3]

This study is particularly interesting because a common notion of the passive is that its relationship to the active voice is defined in terms of subjects and objects. Whether or not this is true in an adult, it appears that this is not the way that children learn this alternation. It seems that children first acquire this systematic alternation in a semantically-limited arena, in which the active-voice patient is promoted to the passive “subject”. Only later do they extend it to a more “semantically abstract” arena in which it is the active-voice object that is promoted to the subject position.

4 A Theoretical Proposal

We wish to test a proposal put forward in Morris [40], which describes an approach to learning grammatical relations without recourse to innate, domain-specific, linguistic knowledge. This model is based on (i) the Goldberg variation of Construction Grammar [24, 25], and (ii) the learning mechanisms of connectionism [50], *inter alia*. The proposal is that the acquisition of grammatical relations occurs as a three-stage process.

In the first stage a child learns verb argument structures as separate, individual “mini-grammars”. This word is used to emphasize that there are no overarching abstractions that link these individual argument structures to other argument structures. Each argument structure is a separate grammar unto itself.

In the second stage the child develops correspondences between the separate mini-grammars; initially the correspondences are based on both semantic and syntactic similarity, later the correspondences are established on purely syntactic criteria. The transition is gradual, with the role that semantics plays decreasing slowly.

For example, the verbs *eat* and *drink* are quite similar to each other, and will “merge” quickly into a larger grammar. Similarly, the verbs *hit* and *kick* will merge early, since their semantics and syntax are similar. While all four of these verbs have agents and patients as verb arguments, there are many semantic differences between the verbs of ingestion and the verbs of physical assault, therefore the merge between these two verb groups will occur later in development.

Ultimately, these agent-patient verbs will merge with experiencer-percept verbs (e.g., *like, fear, see, remember*), percept-experiencer verbs (e.g., *please, frighten, surprise*), and others, yielding a prototypical transitive construction with an extremely ab-

---

[3] The few experiential verbs that they did find in the passive voice in parental input were of the percept-experiencer type (e.g., *frighten, surprise*) rather than the experiencer-percept type (e.g., *fear, like*). Maratsos et al. did not test the children for their comprehension of percept-experiencer verbs.

stract argument structure. The verb-arguments in these abstract argument structures can be identified as “A”, the transitive actor, and “O”, transitive patient (or “object”). In addition there is prototypical intransitive argument structure with a single argument, “S”, the intransitive “subject”. (This schematic description is due to Dixon [19].)

In the third stage, the child begins to associate the abstract arguments of the abstract transitive and intransitive constructions with the coindexing constructions that instantiate the properties of, for example, clause coordination, control structures, and reflexivization. So, for example, an intransitive-to-transitive coindexing construction will associate the S of an intransitive first clause with the deleted co-referent A of a transitive second clause. This will enable the understanding of a sentence like *Max arrived and hugged everyone.* Similarly, a transitive-to-intransitive coindexing construction will associate the A of an initial transitive clause with the S of a following intransitive clause; this will enable the understanding of a sentence like *Max hugged Annie and left.*

Since this association takes place relatively late in the process, necessarily building on layers of abstraction and guided by input, the grammatical relations (of which S, A, and O are the raw material) “grow” naturally into the language-appropriate molds.

From beginning to end this is a usage-based acquisition system. It starts with rote-acquisition of verb-argument structures, and by finding commonalities, it slowly builds levels of abstraction. Through this bottom-up process, it accommodates to the target language. (For other accounts of usage based systems, see also Bybee [14] and Langacker [31, 32, 33].)

5 A connectionist simulation

In this section we present a connectionist simulation to test whether a network could build abstract relationships corresponding to “subjects” and “objects” given an English-like language with a variety of grammatical constructions. This was done in such a way that there is no “innate” knowledge of language in the network. In particular, there are no architectural features that correspond to “syntactic elements”, i.e., no grammatical relations, no features that facilitate word displacement, and so forth. The main assumptions are that the system can process sequential data, and that it is trying to map sequences of words to semantic roles.

The motivation behind the network is the notion that merely the drive to map input words to output semantics is sufficient to induce the necessary internal abstractions to facilitate the mapping. To test this hypothesis, a Simple Recurrent Network [22] was created and tested using the Stuttgart Neural Network Simulator (SNNS). The network is shown in Figure 1.

The network takes in a sequence of patterns representing sentences generated from a grammar. At each time step, a word or end of sentence marker is presented. After each sentence, an input representing “reset” is presented, for which the network is supposed to zero out the outputs. The output patterns represent semantic roles in a slot-based representation. The teaching signal for the roles are given as targets starting from the first presentation of the corresponding filler word, and then held constant throughout the rest of the presentation of the sentence.

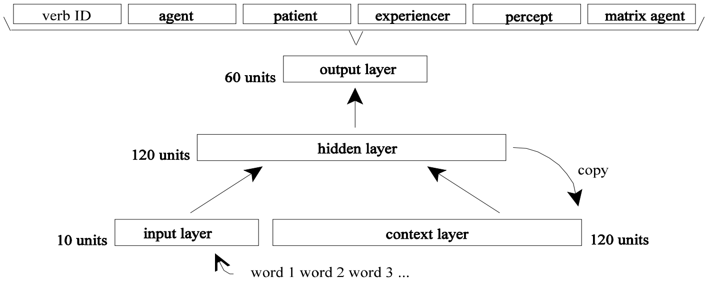

Fig. 1. Network architecture.

The input vocabulary consists of 56 words (plus end of sentence and reset), represented as 10-bit patterns, with 5 bits on and 5 bits off. Of these 56 words, 25 are verbs, 25 are nouns, and remaining 6 are a variety of function words. All of the nouns are proper names. Of the verbs, 5 are unergative (intransitive, with agents as the sole arguments, e.g., *run, sing*), 5 are unaccusative (intransitive, with patient arguments, e.g., *fall, roll*), 10 are "action" transitives (with agent & patient arguments, e.g., *hit, kick, tickle*), and 5 are "experiential" transitives (with experiencer & percept arguments, e.g., *see, like, remember*). In addition there is a "matrix verb", *persuade*, which is used for embedded sentence structures. The 5 remaining words are *who, was, by, and*, and *self.*

The output layer is divided into 6 slots that are 10 units wide. The first slot is the verb identifier, the second through the fifth are the identifiers for the agent, the patient, the experiencer, and the percept. (Note that at most only two of these four slots should be filled at one time.) The sixth slot is the "matrix agent" slot, which will be explained below. The representation of the fillers is unrelated to the representation of the words – the slot fillers only have 2 bits set out of 10. Hence the network cannot just copy the inputs to the slots.

Using the back-propagation learning procedure [49] the network was taught to assign the proper noun identifier(s) to the appropriate role(s) for a number of sentence structures. Thus for the sentence, *Sandy persuaded Kim to kiss Larry,* the matrix agent role is filled by *Sandy,* the agent role is filled by *Kim,* and the patient role is filled by *Larry.* In the sentence, *Who did Larry see,* the experiencer role is filled by *Larry* and the percept role is filled by *who.* Training was conducted for 50 epochs, with 10,000 sentences in each epoch. The learning rate was 0.2, initial weights set within a range of 1.0. There was no momentum.

Examples of the types of sentences and their percentage in the training set are listed below:

1. Simple declarative intransitives (18%). E.g., *Sandy jumped* (agent role) and *Sandy fell* (patient role).

2. Simple declarative transitives (26%). E.g., *Sandy kissed Kim* (agent and patient roles) and *Sandy saw Kim* (experiencer and percept roles).

3. Simple declarative passives (6%). E.g., *Sandy was kissed* (patient role).

4. Questions (20%). E.g., *Who did Sandy kiss?*(agent and patient roles, object is questioned), *Who kissed Sandy?* (agent and patient roles, subject is questioned), *Who*

did Sandy see? (experiencer and percept roles, object is questioned), and *Who saw Sandy?* (experiencer and percept roles, subject is questioned).

5. Control (equi-NP) sentences (25%). E.g., *Sandy persuaded Kim to run* (matrix agent and agent roles), *Sandy persuaded Kim to fall* (matrix agent and patient roles), *Sandy persuaded Kim to kiss Max* (matrix agent, agent, and patient roles) and *Sandy persuaded Kim to see Max* (matrix agent, experiencer, and percept roles).

6. Control (equi-NP) sentences with questions (6%). E.g., *Who did Sandy persuade to run/fall?* (questioning embedded subject, whether agent or patient, of an intransitive verb), *Who persuaded Sandy to run/fall?* (questioning matrix agent; note embedded intransitive verb), *Who persuaded Sandy to kiss/see Max?* (questioning matrix agent; note embedded transitive verb), and *Who did Sandy persuade to kiss Max?* (questioning embedded agent).

The generalization test involved two systematic gaps in the data presented to the network; both involved experiential verbs. The first was passive sentences with experiential verbs e.g., *Sandy was seen by Max.* The second involved questioning embedded subjects in transitive clauses with experiential verbs, e.g., *Who did Sandy persuade to see Max?* Neither of these sentence types occurred with experiential verbs in the training set. The test involved probing these gaps.

The network was not expected to generalize over these two systematic gaps in the same way. The questioning-of-embedded-subject-sentences gap is part of an interlocking group of constructions which “conspire” to compensate for the gap. The “members of the conspiracy” are the transitive sentences (group 2 above), the questions (group 4), and the control sentences (group 5). These sentences are related to each other, and they should cause the network to treat the agents of action verbs and the experiencers of experiential verbs the same. Thus we believe that this gap, which is unattested in parental input, should show some generalization. Our explanation in terms of construction conspiracies would then be the basis for our explanation of many of the overgeneralizations that occur in children.

Meanwhile, the passive gap has no such compensating group of constructions. Only the transitive sentences (group 2) provide support for the passive generalization. This gap corresponds to one that actually exists in parental input. If our model is a good one, we would expect that it should not bridge this gap.

6 Results

In Table 1 we show the result of testing a variety of constructions, some forms of which were trained, and two were not. Five hundred sentences of each listed type were tested. The results were computed using Euclidean distance decisions-each field in the output vector was compared with all possible field values (including the all-zeroes vector), and the fields assigned the nearest possible correct value. For a sentence to be “correct” all of the output fields had to be correct. The two salient lines are for simple passive clauses with experiential verbs, which had a 6.2% success rate, and questioning embedded subjects with experiential verbs, which had a 67.4% success rate. The near complete failure of generalization for simple passive clauses with experiential verbs showed that the nonappearance of experiential verbs in the passive voice in the training set caused

Sentence description	Percent correct
Simple active clauses, action verbs	97.6%
Simple active clauses, experiential verbs	97.6%
Simple passive clauses, action verbs	91.8%
Simple passive clauses, experiential verbs	6.2%
Control (equi-NP) structures	83.6%
Questioning embedded subjects, action verbs	91.4%
Questioning embedded subjects, experiential verbs	67.4%

Table 1. Sentence comprehension using Euclidean distance decisions.

the network to learn the passive voice as a semantically narrow alternation. This is similar to the undergeneralization found by Maratsos et al. [37, 38], discussed above. This gap, as mentioned earlier, has been shown [37] to be one that actually exists in parental input to children.

On the other hand, the questioning of embedded subjects with experiential verbs, which likewise did not appear in the training set, showed much greater generalization, in all likelihood because there is a "conspiracy of syntactic constructions" surrounding this gap. As mentioned above, the simple transitive clauses, questioned simple clauses, and control sentences, were the prime "conspirators".

Simple transitive clauses established the argument structures for both the agent-patient verbs and the experiencer-percept verbs:

- *Roger kissed Susie.* (agent–patient argument structure)
- *Linda saw Pete.* (experiencer–percept argument structure)

Questioned simple clauses established the ability to question the subjects of both argument structures:

- *Who pinched Sandy?* (questioned agent)
- *Who remembered Max?* (questioned experiencer)

Control sentences established embedded clauses for both argument structures:

- *Fred persuaded Ian to tickle Lynn.* (embedded agent–patient. argument structure)
- *Fred persuaded Sam to hate Terry.* (embedded experiencer–percept argument structure)

Questioning embedded agents established the relevant pattern, including the fronting of the embedded, questioned constituent:

- *Who did Raul persuade to tickle Sally?* (embedded questioned agent)

The interlocking patterns above led to extension of this last pattern to experiencer–percept verbs.

The passive gap has no such compensating group of constructions. Only the transitive sentences (group 2) provided support for the passive generalization; as we shall see, these were insufficient to bridge the gap.

Simple transitive clauses established the similarity of argument structures:

- *Sally tickled Jack.* (agent–patient argument structure)

- *Jack liked Sally.* (experiencer–percept argument structure)

Simple intransitive clauses established patients as subjects:

- *Susie fell.* (patient–only argument structure)

Passive sentences, which only occurred with agent–patient verbs, established an alternation between active–voice agent–patient argument structures and passive–voice patient–only argument structures *with the same verbs:*

- *Jack was tickled.* (patient–only argument structure with a verb that is seen in the active voice)

The gap of the questioned–embedded–experiencer was overcome because there was a sufficient number of overlapping constructions and there was a well-established precedent of experiencer subjects. As a result we are seeing a level of abstraction, with the network able to "define", in some sense, the gap in terms of the embedded subject rather than merely an embedded agent.

In order for the gap of the passive-voice for experiencer–percept verbs to be overcome there would have to have been an established precedent of percept–subjects. There were none. There were no percept–only verbs in the data set; indeed, there are arguably no percept–only verbs in English. The gap of the passive-voice for experiencer–percept verbs was not overcome because there was an insufficient number of overlapping constructions, and because there was no precedent of percept–subjects in the data set.

6.1 Analysis of representations in the hidden layer

We wanted to probe the way that the network represented subjects internally, i.e., in the hidden layer. This was done by creating and comparing "subject-variance vectors" for combinations of verb classes and syntactic constructions. Subject–variance vectors are vectors representing the variance of the hidden layer units when only the subject is varied. This should show where the subject is being encoded in the hidden layer. Creating the variance vectors is a three-step process.

To construct these vectors, we presented the network with 25 sentences varying only in their subject. We saved the 120 hidden unit activations at the end of each presentation, and computed the variance on a per unit basis. The variances so computed should then represent "where" the subject is being encoded for that verb/construction combination.

Next we compared the subject-variance vectors within a verb class to each other. An average subject-variance vector was computed for each verb class (for a given construction); this represented the "prototype" subject representation for the verb class.

To test how tightly associated the representations of the subjects of these verb-construction classes were we computed the average Euclidean distance from the prototype to each of the members of the class. For unaccusative (patient-only) and unergative (agent-only) verbs in simple clauses and in embedded clauses the average distances were about 0.5. For transitive verbs, both agent-patient and experiencer-percept verbs, in simple clauses and in embedded clauses, the averages were about 0.3. For passive-voice agent-patient verbs the average was about 0.4. (The fact that intransitive verb-construction combinations have "less disciplined", i.e., less tightly associated, subject

Simple clauses	Other simple clauses	Distance
Transitive Agent	Intransitive Agent	0.69
Transitive Agent	Transitive Experiencer	0.69
Intransitive Agent	Transitive Experiencer	0.82
Transitive Experiencer	Intransitive Patient	1.15
Intransitive Agent	Intransitive Patient	1.18
Transitive Agent	Intransitive Patient	1.40
Simple clauses	Embedded counterparts	Distance
Intransitive Patient	Embedded Intransitive Patient	0.70
Intransitive Agent	Embedded Intransitive Agents	0.73
Transitive Experiencer	Embedded Transitive Experiencer	0.77
Transitive Agent	Embedded Transitive Agent	0.81
Embedded clauses	Other embedded clauses	Distance
Embedded Transitive Agent	Embedded Transitive Experiencer	0.57
Embedded Transitive Agent	Embedded Intransitive Agent	0.67
Embedded Transitive Experiencer	Embedded Intransitive Agent	0.78
Embedded Transitive Experiencer	Embedded Intransitive Patient	1.07
Embedded Intransitive Agent	Embedded Intransitive Patient	1.07
Embedded Transitive Agent	Embedded Intransitive Patient	1.27
Active voice	Passive voice	Distance
Intransitive Patient	Passive Voice Patient	0.72
Transitive Experiential	Passive Voice Patient	1.27
Intransitive Agent	Passive Voice Patient	1.37
Transitive Agent	Passive Voice Patient	1.51
Passive Voice Percept	Passive Voice Patient	0.38

Table 2. Euclidean distances between prototype subject variance vectors.

representations may be explained by the fact that these verbs have only a single verb argument. The network need not “remember” two verb arguments simultaneously; it can therefore be profligate in the manner of the storage of the subject’s identity.)

The third step involved looking at the distances between the prototypes. This allowed us to see how similar prototypes were. The results of our comparisons are shown in Table 2, where we use “Intransitive Agent” for unergative subjects (e.g., *Sandy jumped*) and “Intransitive Patient” for unaccusative subjects (e.g. *Sandy fell*). In general, one can think of distances less than 1.0 as “close” (although none are as close as the within-class distances mentioned above) and distances greater than 1.0 as “far”. With this in mind, Table 2 shows that there are interesting relationships between the instantiations of subjects in various verb-and-construction groups (recall that *all* the entries in the Table correspond to subjects).

First, considering only the simple clauses, we see that the entries divide into two distinct groups. The intransitive patients are relatively far from the other classes, while the agents and experiencers tend to pattern together. To understand why, consider that in transitive constructions with agent-patient verbs, both agents and patients must be present. Therefore the two semantic roles must be stored simultaneously, and thus their representations must be in somewhat different units. Agents, whether transitive or intransitive, will most likely be represented by the same set of units. Note that experiencers *never* need to be stored simultaneously with agents. Therefore their representation can overlap agent-subjects much more than can the representations of patient-subjects. Then the question is why experiencers pattern more with agents than patients. We believe this is because the agent-subjects are simply the most frequent in the training set, and thus have a primacy in “carving out” the location for subjects in the hidden layer. This is also consistent with many linguistic theories where agents are considered the prototypical subjects.

Second, the distances between the matrix clause subjects and their embedded clause counterparts are also close, and in the same range as the distances between “non-antagonistic” subject types.

Third, the embedded clauses essentially replicate the pattern seen in the simple clauses, with agent and experiencer subjects patterning together, and patient subjects at a distance.

Fourth, the passive voice patient subjects are far from active voice subjects, with the (not unexpected) exception of active voice intransitive patients. Clearly, the network has drawn a major distinction between patient-subjects and non-patient subjects. Again, we hypothesize that the network did this simply because of the necessity of storing agents and patients simultaneously.

Finally, we see that passive voice patient subjects are *very* close (within the range of a within-class distance) to passive voice subjects of experiential verbs (percepts). Recall that the network was never trained on experiential verbs in the passive voice and never trained with percept-subjects; the network has basically stored such subjects in the same location as passive voice patient subjects. This is consistent with the failure of the network to correctly process these novel constructions.

We conclude from this analysis of the subject-variance vectors that within a syntactically defined class of verbs, the subjects are stored in very nearly the same set of units. These subject patterns are more similar to each other than they are to the subject patterns for the same class of verbs in other constructions, or to the subject patterns of other classes of verbs. Most importantly, though, the representation of “subject” in the network is controlled by two main factors. First, if the subjects of two sentences must fill the same thematic role, they will be stored similarly. Second, representations are pushed apart according to whether the processing requirements force them to compete for representational resources. In the case of our set of sentence types, the effect is that agents and patients are stored separately because they can appear together, and experiencers are stored very close to agents, since they never appear together. The result is that the instantiation of “subject” in the network amounts to a radial category in the manner of Lakoff’s *Women, Fire, and Dangerous Things* [30]. These relationships are largely in accord with the predictions of the theoretical model sketched out in this paper.

7 Discussion and conclusions

This simulation was intended to demonstrate that the most abstract aspects of language are learnable. There are two broad areas in which this is explored: control of “subjecthood” properties and demonstration of relative abstraction.

In the area of control of properties, this simulation demonstrated that the network was capable of learning to process equi-NP deletion sentences (also known as “control constructions"). This is shown in the ability of the network to correctly process sentences such as *Sandy persuaded Kim to run* (these are shown in groups 5 & 6, in section 5 above). As was seen above, the network was able to correctly understand these sentences at a rate of 84%.

The network's ability to abstract from semantics was shown in the ability of the network to partially bridge the artificial gap in the training set, that of the questioned embedded subject of experiential verbs. The network was able to define the position in that syntactic construction in terms of a semantically-abstract entity, that is, a subject rather than an agent. Consistent with developmental data, the network also did *not* generalize when it should not have. In particular, it did not process passive sentences with perceptual subjects. We have hypothesized that this pattern of generalization and lack of generalization can be explained as a conspiracy of constructions, that bootstrap the processing required for a new construction. Without this scaffolding, the network assimilates the new construction into a known one.

As is clear from the examination of the hidden layer, we can see how the network stores a partially-abstract representation of the subject. We can also see the limitations of abstraction; the network's representation of the subject of a given sentence is also partially specified in semantically loaded units. And, as we have seen in the Maratsos [37] study, this appears to be appropriate to the way that humans learn language. This result is also consistent with Goldberg's theoretical analysis [25] that predicts this semantically-limited scope to certain syntactic constructions.

Of course, we have been preceded by many others in the use of recurrent networks for language comprehension [16, 22, 28, 39, 52]. Most of these previous works impose a great deal of structure on the networks that, in some sense, parallels a preconceived notion of what sentence processing should look like. The previous work to which we owe the greatest debt is that of Elman [22], who developed Simple Recurrent Networks, and St. John & McClelland [52], who applied them to the problem of mapping from sequences of words to semantic representations. There are two main differences between this work and that of St. John & McClelland. In terms of networks, ours is simpler, because we specify in advance an output representation for semantics. While our semantics is simpler, the syntactic constructions used in training are more complex. Indeed, the fact that we focus upon the notion of a grammatical relation and how it could be learned is what differentiates this work from much of the previous work. Such a notion, as shown in the list of characteristic properties, requires a fairly large array of sentence types. Our analysis of the network's representation of this notion also is novel.

One obvious drawback of our work is the impoverished semantics. All of our nouns were glossed as proper names, but they were just simple bit patterns with no inherent structure. The only difference in verb "meanings", aside from a particular bit pattern for a signature, was the set of thematic roles they licensed. A richer semantics would presumably be required to model the earlier stages of the theory, where verbs with similar meanings merge into larger categories. On the bright side, preliminary studies for future work, as well as similar studies by Van Everbroeck [58], indicate that this sort of network can be scaled up in the size of the vocabulary.

In the context of this book, this work demonstrates that a "radical" connectionist approach, that is, one without any additional bells and whistles to force it to be "symbolic", is indeed able to form categories usually reserved for symbolic approaches to linguistic analysis. Indeed, we believe that this sort of approach will eventually show that syntax as a separate entity from semantic processing is an unnecessary assumption. Rather, what we see in our network is that "syntax", in the usual understanding of that term, is part and parcel of the processing required to map from a sequence of input words to a set of semantic roles.

Acknowledgments

The authors would like to thank Adele Goldberg and Ezra Van Everbroeck for their comments and suggestions in the course of this study.

References

[1] N. Akhtar. Characterizing English-speaking children's understanding of SVO word order. to appear.

[2] L. Bloom. *One Word at a Time: The Use of Single Word Utterances Before Syntax*. Mouton de Gruyter, The Hague, 1973.

[3] L. Bloom, K. Lifter, and J. Hafitz. Semantics of verbs and the development of verb inflection in child language. *Language*, 56:386–412, 1980.

[4] H. Borer and K. Wexler. The maturation of syntax. In T. Roeper and E. Williams, editors, *Parameter Setting*. D. Reidel Publishing Company, Dordrecht, 1987.

[5] M. Bowerman. Learning the structure of causative verbs: A study in the relationship of cognitive, semantic and syntactic development. In *Papers and Reports on Child Language Development*, volume 8, pages 142–178. Department of Linguistics, Stanford University, 1974.

[6] M. Bowerman. Semantic factors in the acquisition of rules for word use and sentence construction. In D. M. Morehead and A. E. Morehead, editors, *Normal and deficient child language*. University Park Press, Baltimore, 1976.

[7] M. Bowerman. Evaluating competing linguistic models with language acquisition data: Implications of developmental errors with causative verbs. *Quaderni di semantica*, 3:5–66, 1982.

[8] M. Bowerman. Reorganizational processes in lexical and syntactic development. In E. Wanner and L. R. Gleitman, editors, *Language acquisition: The state of the art*, pages 319–346. Cambridge University Press, Cambridge, 1982.

[9] M. Bowerman. The "no negative evidence" problem: How do children avoid an overgeneral grammar? In John A. Hawkins, editor, *Explaining Language Universals*, pages 73–101. Basil Blackwell, Oxford (UK), 1988.

[10] M. Bowerman. Mapping thematic roles onto syntactic functions: Are children helped by innate linking rules? *Linguistics*, 28:1253–1289, 1990.

[11] M. D. S. Braine. On learning the grammatical order of words. *Psychological Review*, 70:323–348, 1963.

[12] M. D. S. Braine. *Children's first word combinations*, volume 41 of *Monographs of the Society for Research in Child Development*. University of Chicago Press, Chicago, 1976.

[13] R. Brown. *A First Language: The Early Stages*. Harvard University Press, Cambridge MA, 1973.

[14] J. Bybee. *Morphology: A Study of the Relation between Meaning and Form*. John Benjamins, Amsterdam, 1985.

[15] E. V. Clark. Early verbs, event-types, and inflections. In C. E. Johnson and J. H. V. Gilbert, editors, *Children's Language*, volume 9, pages 61–73. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

[16] Garrison W. Cottrell. *A Connectionist Approach to Word Sense Disambiguation*. Research Notes in Artificial Intelligence. Morgan Kaufmann, San Mateo, 1989.

[17] J. G. de Villiers, M. Phinney, and A. Avery. Understanding passives with non-action verbs. Paper presented at the Seventh Annual Boston University Conference on Language Development, October 8-10, 1982.

[18] R. M. W. Dixon. *The Dyirbal Language of North Queensland*. Cambridge University Press, Cambridge UK, 1972.

[19] R. M. W. Dixon. Ergativity. *Language*, 55:59–138, 1979.

[20] R. M. W. Dixon. *Ergativity*. Cambridge University Press, Cambridge, 1994.

[21] D. Dowty. Thematic proto-roles and argument selection. *Language*, 67(3):547–619, 1991.

[22] J. L. Elman. Finding structure in time. *Cognitive Science*, 14:179–211, 1990.

[23] W. A. Foley and R. D. Van Valin Jr. *Functional Syntax and Universal Grammar*. Cambridge University Press, Cambridge, 1984.

[24] A. E. Goldberg. *Argument Structure Constructions*. PhD thesis, University of California, Berkeley, 1992.

[25] A. E. Goldberg. *A Construction Grammar Approach to Argument Structure*. University of Chicago Press, Chicago, 1995.

[26] K. Hirsh-Pasek and R. M. Golinkoff. *The Origins of Grammar: Evidence from Early Language Comprehension*. MIT Press, Cambridge MA, 1996.

[27] N. M. Hyams. *Language acquisition and the theory of parameters*. D. Reidel Publishing Company, Dordrecht, 1986.

[28] Ajay N. Jain. A connectionist architecture for sequential symbolic domains. Technical Report CMU-CS-89-187, Carnegie Mellon University, 1989.

[29] E. L. Keenan. Towards a universal definition of "Subject". In C. Li, editor, *Subject and Topic*, pages 303–334. Academic Press, New York, 1976.

[30] G. Lakoff. *Women, Fire, and Dangerous Things*. Chicago University Press, Chicago, 1987.

[31] R. W. Langacker. *Theoretical Prerequisites*, volume 1 of *Foundations of Cognitive Grammar*. Stanford University Press, Stanford, 1987.

[32] Ronald W. Langacker. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Mouton de Gruyter, Berlin, 1991.

[33] Ronald W. Langacker. *Descriptive Application*, volume 2 of *Foundations of Cognitive Grammar*. Stanford University Press, Stanford, 1991.

[34] B. Levin. *On the Nature of Ergativity*. PhD thesis, MIT, 1983.

[35] C. D. Manning. *Ergativity: Argument Structure and Grammatical Relations*. PhD thesis, Stanford, 1994.

[36] A. P. Marantz. *On the nature of grammatical relations*. MIT Press, Cambridge MA, 1984.

[37] M. Maratsos, D. E. C. Fox, J. A. Becker, and M. A. Chalkley. Semantic restrictions on children's passives. *Cognition*, 19:167–191, 1985.

[38] M. Maratsos, S. A. Kuczaj II, D. E. C. Fox, and M. A. Chalkley. Some empirical studies in the acquisition of transformational relations: Passives, negatives, and the past tense. In W. A. Collins, editor, *Children's Language and Communication: The Minnesota Symposia on Child Psychology*, volume 12, pages 1–45. Lawrence Erlbaum Associates, Hillsdale NJ, 1979.

[39] Risto Miikkulainen. *Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory*. MIT Press, Cambridge, MA, 1993.

[40] W. C. Morris. *Emergent Grammatical Relations: An Inductive Learning System*. PhD thesis, University of California, San Diego, 1998.

[41] L. Naigles. Children use syntax to learn verb meanings. *Journal of Child Language*, 17:357–374, 1990.

[42] L. Naigles, K. Hirsh-Pasek, R. Golinkoff, L. R. Gleitman, and H. Gleitman. From linguistic form to meaning: Evidence for syntactic bootstrapping in the two-year-old. Paper presented at the Twelfth Annual Boston University Child Language Conference, Boston MA, 1987.

[43] R. Olguin and M. Tomasello. Twenty-five-month-old children do not have a grammatical category of verb. *Cognitive Development*, 8:245–272, 1993.

[44] J. M. Pine, E. V. M. Lieven, and C. F. Rowland. Comparing different models of the development of the english verb category. MS, Forthcoming.

[45] J. M. Pine and H. Martindale. Syntactic categories in the speech of young children: The case of the determiner. *Journal of Child Language*, 23:369–395, 1996.

[46] S. Pinker. *Language Learnability and Language Development*. Harvard University Press, Cambridge MA, 1984.

[47] S. Pinker. *Learnability and Cognition: The Acquisition of Argument Structure*. MIT Press, Cambridge MA, 1989.

[48] S. Pinker, D. S. LeBeaux, and L. A. Frost. Productivity and constraints in the acquisition of the passive. *Cognition*, 26:195–267, 1987.

[49] D. E. Rumelhart and J. L. McClelland. On learning the past tenses of English verbs. In J. L. McClelland and D. E. Rumelhart, editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 2, pages 216–271. The MIT Press, Cambridge MA, 1986.

[50] D. E. Rumelhart and J. L. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1. The MIT Press, Cambridge MA, 1986.

[51] P. Schachter. The subject in Philippine languages: Topic, Actor, Actor-Topic, or none of the above? In C. Li, editor, *Subject and Topic*, pages 491–518. Academic Press, New York, 1976.

[52] Mark F. St. John and James L. McClelland. Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, 46:217–257, 1990.

[53] M. Tomasello. *First verbs: A case study of early grammatical development*. Cambridge University Press, Cambridge, 1992.

[54] M. Tomasello and P. J. Brooks. Early syntactic development: A construction grammar approach. In M. Barrett, editor, *The Development of Language*. UCL Press, London, in press.

[55] M. Tomasello and R. Olguin. Twenty-three-month-old children do have a grammatical category of noun. *Cognitive Development*, 8:451–464, 1993.

[56] V. Valian. Syntactic categories in the speech of young children. *Developmental Psychology*, 22:562–579, 1986.

[57] V. Valian. Syntactic subjects in the early speech of American and Italian children. *Cognition*, 40:21–81, 1991.

[58] E. Van Everbroeck. Language type frequency and learnability: A connectionist appraisal. In M. Hahn and S. C. Stoness, editors, *The Proceedings of the Twenty First Annual Conference of the Cognitive Science Society*, pages 755–760, Mahwah NJ, 1999. Lawrence Erlbaum Associates.