Harmonic Grammar - A formal multi-level connectionist theory of linguistic well-formedness: An application

Géraldine Legendre[1,2], Yoshiro Miyata[1,3,4], Paul Smolensky[1,3,4]

[1]*Institute of Cognitive Science,* [2]*Department of Linguistics,* [3]*Department of Computer Science,* & [4]*Optoelectronic Computing Systems Center, University of Colorado at Boulder*

Abstract

We describe *harmonic grammar*, a connectionist-based approach to formal theories of linguistic well-formedness. The general approach can be applied to various kinds of linguistic well-formedness, e.g., phonological and syntactic. Here, we address a syntactic problem: unaccusativity. Harmonic grammar is a two-level theory, involving a distributed, lower level connectionist network whose relevant aggregate computational behavior is described by a local, higher level network. The central hypothesis is that the connectionist well-formedness measure called "harmony"[1] can be used to model linguistic well-formedness; what is crucial about the relation between the lower and higher level networks is that there is a harmony-preserving mapping between them: they are *isoharmonic* (at least approximately). A companion paper (Legendre, Miyata, & Smolensky, 1990; henceforth "LMS_2") describes the theoretical basis for the two level approach, starting from general connectionist principles. In this paper, we discuss the problem of unaccusativity, give a high level characterization of harmonic syntax, and present a higher level network to account for unaccusativity data in French. We interpret this network as a fragment of the grammar and lexicon of French expressed in "soft rules." Of the 760 sentence types represented in our data, the network correctly predicts the acceptability in all but two cases. This coverage of real, problematic syntactic data greatly exceeds that of any other formal account of unaccusativity of which we are aware.

The Linguistic Problem: Unaccusativity

Since the goal of this work is to develop a genuine refinement of current symbolic linguistic accounts, we begin by devoting a few pages to a discussion of a linguistic problem and current symbolic approaches to it. Verbs have traditionally been characterized as either transitive or intransitive. Following a Relational Grammar proposal put forward by Perlmutter (1978) and adopted since by other grammar frameworks, e.g., Burzio (1986) in Government-Binding theory, formal syntacticians today generally agree that it is necessary to further divide intransitive verbs into two types, *unergatives* and *unaccusatives*, each associated with a different syntactic configuration. In Government-Binding theory, for example, the three classes of verbs differ with respect to their d(eep)-structure, as illustrated in (1).

(1) a. Transitive Verb: $[_S$ NP $[_{VP}$ V NP]]
b. Unergative Intransitive Verb: $[_S$ NP $[_{VP}$ V]]
c. Unaccusative Intransitive Verb: $[_S$ $[_{VP}$ V NP]]

While transitive verbs take two NP arguments, a subject and a direct object, unergative and unaccusative take only one: a deep subject (VP-external argument) in the case of unergatives, a deep direct object (VP-internal argument) in the case of unaccusative verbs.

The unaccusative/unergative distinction is motivated by the cross-linguistic fact that in certain syntactic contexts in which a verb may be embedded, one class of intransitive verbs are acceptable while the remaining are not. These syntactic environments are called *diagnostic contexts for unaccusativity*; we'll call them "diagnostic contexts" or simply "contexts." An example from French, Object Raising, is illustrated in (2).

(2) a. **La vérité** est facile à faire ***dire*** aux enfants.
The truth *is easy to make children* ***tell.***
b. ***Les enfants** sont faciles à faire ***dire*** la vérité.
The children *are easy to make tell the truth.*
c. **La glace** est facile à faire ***fondre.***
Ice *is easy to make melt.*
d. ***Les étudiants** sont faciles à faire ***travailler.***
The students *are easy to make work.*

Each diagnostic context can be viewed as a sentence frame with two slots, typographically identified in (2): an **argument slot** for an NP, which is an argument of the verb filling the *predicate slot*. French (but not English) Object Raising exemplifies the following characteristic property of diagnostic contexts: when a transitive verb is inserted in the predicate slot, the acceptability of the sentence depends on whether the argument slot is filled by the deep subject or the deep direct object of the verb. In (2a), an acceptable sentence results when the deep direct object of *dire* "tell" — **la vérité** "truth" — appears in the argument slot, but in (2b) an unacceptable sentence [marked by "*"] results when the deep subject appears instead — **les enfants** "children". Paralleling this contrast between subject and direct object is a contrast in the behavior of two classes of intransitive verbs. The sole argument of ***fondre*** "melt" is acceptable in the argument slot (2c), while in (2d) the argument of ***travailler*** "work" is not. Since the argument of ***fondre*** is acceptable, like the deep direct object of ***dire***, it is classified as unaccusative; since the argument of ***travailler*** is unacceptable, like the deep subject of ***dire***, it is classified as unergative.

In English, it turns out, diagnostic contexts for unaccusativity are few in number,[2] but this is by no means the case cross-linguistically. The subject of much current research in syntax and semantics, unaccusativity contexts are cross-linguistically a rich source of interesting patterns. Of central concern is the problem of *unaccusativity mismatches*, which are of two general types. First, across languages, synonymous verbs may be unaccusative in some languages and unergative in others. What is relevant to this paper, however, are the language-internal mismatches: the existence of many intransitive verbs that behave unaccusatively in some contexts and unergatively in others.

A good language in which to study language-internal unaccusativity mismatches is French. Legendre (1989, forthcoming) has identified 10 diagnostic contexts in French, and argued that a necessary and sufficient condition for identifying a verb as unaccusative is that it behave unaccusatively in at least one of these contexts. These contexts display a highly complex pattern of mismatches. It is a subset of these data for which we provide a new type of formal account in this paper. For this initial study, we selected four of these 10 diagnostic contexts; these contexts identify the largest numbers of unaccusative verbs. These four contexts are indicated in Table 1; the first is Object Raising, illustrated in (2).

Table 1.

Context	Example	Translation
Object Raising (OR)	**La glace** est facile à faire ***fondre.***	**Ice** *is easy to make* ***melt.***
Croire "believe" (CR)	Je croyais **Marie** déjà *sortie.*	***I believed*** **Marie** *to have already* ***gone out.***
Participial Absolute (PA)	*Parti* avant l'aube, **Pierre** est arrivé à destination le jour même.	*Gone before dawn,* **Pierre** *arrived to destination on the same day.*
Reduced Relatives (RR)	**La neige** *fondue* a formé de la boue.	*The melted* snow *formed mud.*

In current linguistic theory, the contrasts between unaccusative and unergative intransitive verbs in diagnostic contexts are accounted for by some linguists through syntactic means and by others through semantic means. Syntactic accounts appeal to (1) structural differences between unergatives and unaccusatives, like those posited in (1b,c), (2) structural similarities between the unergative argument and the subject of transitives (1b,a), (3) structural similarities between the unaccusative argument and the direct object of transitives (1c,a), (4) structural restrictions on allowed arguments in diagnostic contexts, and (5) lexical markings indicating which intransitives are unaccusative and which are unergative. Earlier studies led to the claim that the class of unaccusative verbs across languages is not uniform semantically, but uniform syntactically, as expressed in (1c) (Rosen, 1984); however, language-internal mismatches cannot be accounted for on syntactic grounds alone (Legendre, 1989, forthcoming).

Semantic accounts appeal to (1) semantic differences among arguments along various feature dimensions; (2) semantic and aspectual[3] differences among predicates along various feature dimensions; and (3) restrictions in diagnostic contexts on allowed semantic features of arguments and predicates. Semantic features of arguments that are cross-linguistically relevant for unaccusativity include volitionality and animacy, while relevant semantic/aspectual features of predicates include telicity and progressivity,[4] and the classification of verbs as involving activity, accomplishment, achievement, or state (Van Valin, to appear).

Focussing on mismatches, several recent studies (e.g., Zaenen, 1989; Van Valin, to appear) have argued that unaccusativity phenomena can be accounted for on purely semantic grounds by assuming that each diagnostic context involves a simple semantic restriction such as a constraint on the value of some semantic feature. However, this does not appear to be the case for the French data addressed here: at least, the simple semantic restrictions that have been proposed for other languages clearly do not work (Legendre, forthcoming). On the other hand, while they are not sufficient by themselves to provide simple semantic *rules* to account for the data, the argument features volitionality (VO) and animacy (AN), and the predicate features telicity (TE) and progressivity (PR), show strong *tendencies* to influence acceptability in diagnostic contexts (Legendre, forthcoming).

Accounting for the French data appears to require combining syntactic and semantic restrictions in a complex way. This is just what is offered by the connectionist account described here. Replacing the all-or-none categories and rules used in traditional symbolic accounts with formalizations of the semantic and syntactic *tendencies* in the data — *soft rules* — allows our account to achieve a degree of coverage of unaccusativity data, including a complex pattern of mismatches, that, to our knowledge, is unparalleled in the existing literature. At the same time, the new approach naturally addresses two aspects of this problem which are quite difficult to naturally capture in a symbolic approach.

(3) a. ***The graded character of the unaccusative/unergative categories.*** Some intransitive French verbs behave unaccusatively in six of the ten contexts, others four, and still others only one.[5] Clearly the condition of Legendre (1989) — an intransitive verb is unaccusative if and only if it behaves unaccusatively in at least one diagnostic context — is failing to capture the fact that "some verbs are more unaccusative than others," an aspect of the phenomenon that is formally expressed and exploited in our connectionist account.

b. ***The graded character of the acceptability judgements in diagnostic contexts.*** Our account formally predicts not only the polarity of acceptability judgements, but also their strength. While we admit that the accuracy of our account in predicting judgement strengths can stand improvement, it is nonetheless a virtue of our connectionist approach that it makes precise, falsifiable predictions of gradations of acceptability.

Harmonic grammar

The goal of harmonic grammar is to provide a framework, derived from basic connectionist principles, in which regularities in linguistic well-formedness are expressed as *tendencies, preferences* or *soft rules*, rather than as hard rules. This framework is to be a *formal* one in the sense that soft rules are to be specified with sufficient precision to permit precise, falsifiable predictions of the acceptability or well-formedness of sentences.

In its most intuitive form, a central idea of harmonic grammar is to replace hard rules or constraints on well-formedness of the form (4a) with the corresponding soft rule or constraint in (4b).

(4) a. Condition X must never be violated in well-formed structures.

b. If Condition X is violated, then the well-formedness of the structure is diminished by C_X.

For the unaccusativity problem, examples of relevant grammatical and lexical rules are shown in (5) and (6), respectively.

(5) a. In the Object Raising context, the argument can never be a deep subject in a well-formed sentence.

b. In the Object Raising context, if the argument is a deep subject, then the well-formedness of the sentence is diminished by $C_{OR,Subj}$.

(6) a. The argument of *fondre* must not be a deep subject.

b. If the argument of *fondre* is a deep subject, then the well-formedness of the sentence is diminished by $C_{fondre,Subj}$.

Obviously, in harmonic grammar, well-formedness is quantitative; in fact, well-formedness corresponds to the connectionist quantity called *harmony*, and central among the connectionist principles from which the formalism of harmonic grammar is derived are those relating to harmony. The full description of the harmonic grammar formalism, and its derivation from basic connectionist principles, requires combining the description presented here with that found in LMS_2.

The presence of the numerical constants C_X in the soft rules of harmonic grammar would make the formalism hopelessly unwieldy were it not for the fact, discussed below, that these numbers can be computed automatically by presenting the linguistic data to an appropriately designed connectionist network using an appropriate learning rule. (This particular learning procedure is, however, not viable as a model of language acquisition; among other things, it relies on both positive and negative evidence.)

The methodology of harmonic grammar can be described in general terms as follows:

(7) a. For a given problem concerning the well-formedness of certain linguistic structures, choose a rule-based approach (or approaches) to the problem — corresponding to (4a) — and a target set of data that exemplify the problem.

b. Based on the selected rules, identify a set of relevant features for describing the structures appearing in the data.

c. Based on the selected rules, identify which features can plausibly enter together into well-formedness constraints.

d. Embody these constraints as connections in a network. This network takes as input a featural description of a structure, and produces as output a graded well-formedness or acceptability measure.

e. Train the network on the target data. The resulting connection weights can be interpreted as the constants C_X in the soft rules of (4b).

f. Analyze the network, extracting from it accounts of the general patterns in the target data that provide new linguistic insights into the original problem.

In the preceding section, we addressed (7a,b) for the specific application of unaccusativity in French. For (7a), we identified rule-based accounts based on structural considerations and on semantic/aspectual considerations, and identified four diagnostic contexts for unaccusativity. For this first research effort, we treat each diagnostic context as a unit, without making use of its internal structure, which is usually quite complex (by connectionist standards). The actual target data is displayed in Table 2. The first four columns correspond to the four contexts, and each row corresponds to argument/predicate pairs with a given set of features that give rise to a certain pattern of acceptability across the contexts. Each entry in the table indicates the acceptability of a given row's argument/predicate pair in a given column's syntactic context. The symbols "+", "+?", "?", "-?", and "-" indicate "acceptable," "marginally acceptable," "of indeterminate acceptability," "marginally unacceptable," and "unacceptable," respectively. The judgements are those of several mutually consistent informants (one of whom is Legendre). Note that these data are what Chomsky (1965) would consider linguistic *performance*, rather than *competence* — *acceptability* rather than *grammaticality* — since they are raw intuitive acceptability judgements from informants rather than outputs from an underlying grammar hypothesized by the linguist and presumed to operate free of obscuring performance factors.

Table 2

OR	CR	PA	RR	Feature	Predicate	OR	CR	PA	RR	Feature	Predicate	OR	CR	PA	RR	Feature	Predicate
-	-	-	-	- - - +	sévir (4)	-?	+	+	+	- - - +	se retirer	+?	+	+	+	+ + +? +?	se recroqueviller
-	-	-	-	- - + -	ouvrir sur	-?	+	+	+	- - + +	s'ouvrir	+?	+	+	+	+ + + ?	sortir
-	-	-	-	- + - -	durer (4)	-?	+	+	+	- + - +	geler	+?	+	+	+	+ + + +	s'exiler
-	-	-	-	- + - -?	baisser	-?	+	+	+	- + + -	parvenir	+	-	-	-	- - - +	souffler (7)
-	-	-	-	- + - +	souffrir (2)	-?	+	+	+	+ + - +	se retirer	+	-	-	-	- + - +	tousser (4)
-	-	-	-	+ - - +	marcher	?	-	-	-	+ + - -	éternuer	+	-	-	-	+ + - +	se taire (3)
-	-	-	-	+ + - -?	déambuler	?	-	-	-	+ + + +	reculer	+	-	-	-	+ + + +	réagir (2)
-	-	-	-	+ + - ?	exagérer	?	+	+	+	- + - +	vieillir	+	+?	+?	+?	- - - +	se modifier
-	-	-	-	+ + - -	être (10)	?	+	+	+	- + + -	naître	+	+?	+?	+?	- - + +	se dissiper
-	-	-	-	+ + - +	agir (29)	?	+	+	+	- + + +	guérir	+	+?	+	+?	- - - +	peler
-	-	-	-	+ + + -	aller	?	+	+	+	+ + - +	monter (2)	+	+?	+	+	- - - +	vieillir
-	-	-	-	+ + + -?	courir	?	+	+	+	+ + + -	entrer (3)	+	+	-	+?	- - + +	pâlir
-	-	-	-	+ + + +	sauter (2)	?	+	+	+	+ + + +	partir	+	+	+?	+?	- - - +	se disperser
-	-	-	+	- - - -	rester	+?	-	-	-	- - - +	durer	+	+	+?	+?	- + + +	changer
-	?	+?	?	+ + - -	tomber	+?	-	-	-	+ + + ?	capituler	+	+	+?	+	- - - +	pousser
-	+	+	+	- + - -	décéder	+?	-	-	-	- + - +?	rougir (2)	+	+	+?	+	- - + -	éclater
-	+	+	+	- + + -	périr	+?	-	-	-	+ + + -	trébucher	+	+	+	+	- - - +?	jaillir (2)
-	+	+	+	+ + - -	s'effacer (2)	+?	+?	+	+?	- - + +	s'éloigner	+	+	+	+	- - - +	tomber (23)
-	+	+	+	+ + - -?	se réfugier	+?	+?	+	+?	+ + - +	s'éloigner	+	+	+	+	- - + -	entrer (2)
-	+	+	+	+ + + -	surgir	+?	+	+?	+	- - - +	surgir	+	+	+	+	- - + ?	sauter
-?	-	-	-	- - - ?	subsister	+?	+	+	+	- - - +	sortir	+	+	+	+	- - + +?	arriver
-?	-	-	-	- - - +	empirer	+?	+	+	+	- - + -	venir	+	+	+	+	- - + +	s'éteindre (12)
-?	-	-	-	+ + - +	sourire (3)	+?	+	+	+	- - + +	s'effacer	+	+	+	+	- + - +	grossir
-?	-	-	-	+ + + +	s'esquiver (3)	+?	+	+	+	- + + -	tomber	+	+	+	+	- + + +	s'évanouir (2)
-?	+?	-	+?	- + - -?	peler	+?	+	+	+	+ + - -	disparaître	+	+	+	+	+ + - +	se réunir (3)
-?	+?	+	+	+ + + +	venir	+?	+	+	+	+ + - ?	se dissimuler	+	+	+	+	+ + + +	se noyer (4)
-?	+	-	+?	? + + +	pâlir	+?	+	+	+	+ + - +	sombrer						
-?	+	-	+	- + - +	grandir	+?	+	+	+	+ + +? ?	se blottir						

For (7b), we identified the structural argument features *deep subject/direct object*, (which we abbreviate "1"/"2"), the argument semantic features *volitional* ("VO") and *animate* ("AN"), and the predicate aspectual features *telic* ("TE") and *progressive* ("PR"). In addition, each individual intransitive verb is individuated for the purpose of enabling its lexical entry to indicate its particular preference for its argument being a 1 or a 2 (as in the syntactic account). In Table 2, at the right of each row is the identity of the predicate preceded by the values of the four semantic/aspectual features (in the order VO, AN, TE, PR). If there are more than one argument/predicate pairs in the category, the identity of one of such predicates is shown followed by the number of pairs in the category in parentheses. The rows are sorted, essentially going from most unergative to most unaccusative behavior. Table 2 contains 190 rows (argument/predicate pairs) involving 143 different predicates (intransitive verbs). This constitutes acceptability judgements for 760 different sentence types.

In the rest of this paper, we will discuss selected aspects of steps (7c,d,f), and describe step (7e). Most of LMS_2 addresses the crucial steps (7c,d) further, while further development of (7f) is the goal of current research.

The connectionist network

Having set up the linguistic basis for the connectionist account, steps (7a,b), we now describe the connectionist network, illustrated in Figure 1. We begin with a partial description of steps (7c,d), which omits a number of considerations taken up in LMS_2 which depend upon general connectionist principles concerning the distributed representation of structures, the harmony function, and its role in connectionist processing. LMS_2 shows how the network we now discuss is actually mathematically derived as an approximate higher level description of a lower level model based on these connectionist principles, together with the constraints imposed by the elements of linguistic analysis selected in steps (7a-c). This mathematical derivation is responsible for the many aspects of this network which would otherwise seem to be totally ad hoc stipulations.

Input sentences

Each of the 760 input sentences of Table 2 is first coded using a set of variables which are divided into the following four groups:

(8) a. *Context* : the four diagnostic contexts (OR, CR, PA, and RR) are represented by four variables, one for each context. For a given sentence, the variable corresponding to the appropriate context is given the value 1, and the others 0.

b. $A_{feature}$: two variables for the two argument features, VO and AN, each having a value between 1 (+) and 0 (–).

c. $P_{feature}$: two variables for the two predicate features TE and PR, each having a value between 1 (+) and 0 (–).

d. $P_{individual}$: 143 variables each representing one of the 143 predicates. For a given sentence, the variable corresponding to the appropriate predicate is given the value 1, and the others 0.

In the connectionist network, there is a separate input unit for each of these 151 variables (143 identifying individual predicates + 8 others). These units are indicated in Figure 1 by squares. In addition, there is a single output unit whose activity value indicates the network's prediction of the acceptability of the input sentence.

The architecture

The structure of the network is designed to ensure that: (1) the number of free parameters (connection weights) in the network is small compared to the number of data points (760) used to determine them; and (2) each parameter has a clear linguistic interpretation. The following linguistic considerations constrain the architecture of the network:

(9) a. The network should be able to represent the crucial syntactic distinction deep subject ("1") vs. direct object ("2"), which we will call the "structural feature." Since the structural feature is not explicitly present in the input sentence, the network is given two hidden units to use, one representing feature value 1 and the other value 2.

b. For a given context like OR, the grammar can include well-formedness constraints on the structural feature (e.g., "OR prefers 2," as in (6a)) and on the *features* of the predicate (e.g., "OR prefers +PR") *but not* the *identity* of the individual predicate (*not* "OR prefers the predicate *fondre* and prohibits the predicate *éternuer*").

c. For a given context, the grammar can include well-formedness constraints on $A_{feature}$, the argument features (e.g., "OR prefers -VO").

d. A given context can directly bias well-formedness.[6]

e. $A_{feature}$ and $P_{feature}$ can have well-formedness constraints in conjunction with each other (e.g., "+TE prefers +VO") but individually neither one of them can directly constrain well-formedness (e.g., *not* "+AN is absolutely preferred to -AN").

f. $A_{feature}$ and $P_{feature}$ can have preferences for the structural feature (e.g., "+VO prefers 1," "+TE prefers 2").

g. Each individual predicate $P_{individual}$ can have its own preference for the structural feature (a lexical marking corresponding to unaccusative/unergative, e.g. (6b)).
h. Each individual predicate cannot have an absolute bias on grammaticality.

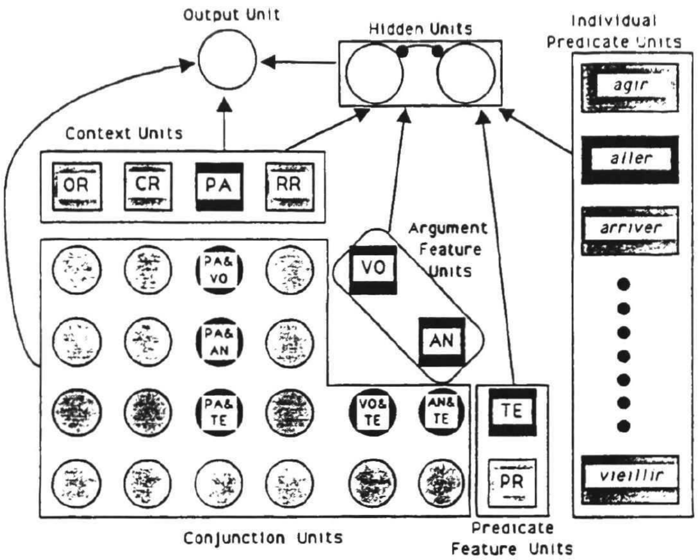

Figure 1. The network architecture.

Together with the connectionist considerations discussed in LMS_2, these linguistic considerations lead to the architecture shown in figure 1. The Conjunction Units represent the constraints that are expressed in terms of conjunctions between pairs of input variables. Each Conjunction Unit is activated in proportion to the product of a particular pair of input units (e.g., "OR's preference for VO" is implemented through a conjunctive unit whose activation is the product of that of the OR unit and that of the VO unit). $P_{individual}$ is not involved in these conjunctions because of the constraint (9b) above.

The Conjunction Units are directly connected to the output unit, following (9b,c,e), as are the Context Units (9d). The Context Units are also connected to the hidden units (9b), as are the Predicate Feature Units, Argument Feature Units, (9f) and Individual Predicate Units (9g). The two hidden units are linear units (their activations are the same as the net inputs). Since the two hidden units are presumed to represent two mutually exclusive structures, they compete against each other, and only the winner of the two (with higher activation) is allowed to send its activation to the output unit. Each connection weight to hidden unit 1 is constrained to be the negative of the corresponding weight to hidden unit 2.[7]

Thus, the net input to the output unit O_{net} is computed as the sum of the contribution of the conjunctions, the contribution of Context Units, and the contribution from the winning hidden unit. As shown in LMS_2, by construction, this net input is equivalent to the global harmony value of the lower level network when it represents the input sentence, which is the network's estimate of the well-formedness of the input sentence. The actual activation value of the output unit, which represents the acceptability estimate, is assumed to be a monotonically increasing function of this net input or global harmony. In the simulation, a logistic sigmoid function $1/(1+e^{-x})$ is used, as is now common in connectionist networks.[8]

The total number of independent connection strengths in the network is 175; 143 embody the lexical preferences of individual verbs for the structural feature, while the remaining 32 are general grammatical preferences.

Training

The network was trained using the informants' acceptability judgements for the 760 sentences shown in Table 2 as the target activation values for its output unit. The standard back propagation algorithm (Rumelhart, Hinton, Williams, 1986) was used to adjust the connection weights which performed a gradient descent (with momentum) on the squared error between the output and the target averaged across all the sentences. For the weights to the hidden units, the winning hidden unit was first allowed to change its in-coming weights, and the weights to the other hidden unit were then constrained to be the negative of the winner's weights. The direct connections to the output unit were adjusted in the normal way.

Results

As the performance measure, we counted two kinds of errors the network made after training. The network made a *major error* when its output and the target were in the opposite sides of 0.5 (e.g., the output was 0.7 when the target was 0.3). It made a *minor error* when the output differed from the target by more than 0.1. Since the metric of the target values was somewhat arbitrary, we have initially been more concerned with the number of major errors the network made than the number of minor errors. The network tended to be quite close to a solution in less than 1000 iterations through the training sentences and the number of major and minor errors changed little afterwards, but the training was continued until 5000 iterations.

The network's performance varied slightly depending on a few learning parameters. The best network made as few as 2 major errors and 64 minor errors out of the 760 sentences. The worst case was 9 major and 120 minor errors. Despite these variations, the following observations showed consistencies in the solutions. First, the errors were very consistent across solutions. In fact, there was a hierarchy of (major) errors such that if network A made less errors than network B, the errors made by A were always a subset of errors made by B. Second, it is possible to predict what acceptability patterns across contexts are possible for a new predicate/argument pair with given features by varying the connection weight for the predicate. (For example, it might be predicted that only four patterns of acceptability are possible for any predicate/argument pair with the features -VO, +AN, +TE, and +PR out of 2^4 possible patterns.) The networks were fairly consistent in their predicted patterns of acceptability. Third, the network always seemed to solve this task by trying to strike a delicate balance among its weights. For example, it was typical that in order to turn on the output unit by generating +1.5 at the net-input, the network used 8 positive contributions adding up to 32.5 and 7 negative contributions adding up to -31.0. As a result, a decision could be flipped by only a slight change to any of the contributing weights.

Conclusion

We have presented a connectionist network that can be interpreted as a partial grammar and lexicon of French containing 175 soft rules (143 in the lexicon, 32 in the grammar) refering to both structural and semantic features. This network makes precise predictions of the acceptability of a particular range of sentences, and is able to account for the complex pattern of unaccusative mismatches exhibited in 760 sentence types, making only two major errors. The numerical strengths of the soft rules are computed from the data by a connectionist learning algorithm. We are currently in the process of analyzing the network to understand its means of accounting for the regularities in the phenomenon, in the hopes of gaining new linguistic insights. If this analysis proves successful, we believe we will have demonstrated that a new kind of linguistic formalism can provide the current best account of a challenging problem combining syntax and semantics.

This new formalism, harmonic grammar, has a number of advantages. On the one hand, the formalism has several of the attractive features of formal symbolic linguistics; it allows us to:

(10) a. do formal analysis: hypothesize a set of grammatical principles that are sufficient to make definite, falsifiable predictions
b. do explanatory analysis: work within a constrained set of possibilities, and go beyond mere description of the data
c. study constraints on linguistic well-formedness
d. incorporate analytical insights from existing formal linguistics

On the other hand, it has certain characteristics that are often felt to be lacking in formal symbolic linguistics; it allows us to:

(11) a. use data on linguistic performance rather than competence
b. deal with the full complexity of data in a specific area of language
c. work within a formalism that is mathematically derived from connectionism, a cognitively (and statistically) based computational framework; we are in a good position from which to pursue integration with other cognitive language models (e.g., acquisition, real-time processing, and neurolinguistic models)

Acknowledgements

Warm thanks to Alan Prince, for very helpful discussions, and especially, a great pot of chicken soup and the term "isoharmonic." This work owes its existence to Mike Mozer, who failed to convince us not to do it. Thanks also to Jim Martin for his valuable comments on an earlier version. This work has been supported by NSF grants IRI-8609599 and ECE-8617947 to PS, by a grant to PS from the Sloan Foundation's computational neuroscience program. PS (in part) and YM have also been supported by the Optical Connectionist Machine Program of the Center for Optoelectronic Computing Systems, which is sponsored in part by NSF/ERC grant CDR-8622236 and by the Colorado Advanced Technology Institute, an agency of the State of Colorado. GL has been supported in part by a Junior Faculty Development Award from the Council on Research and Creative Work, University of Colorado, Boulder. The authors are listed in alphabetical order.

Footnotes

1. Other connectionist approaches to linguistics appealing to the notion of harmony include John Goldsmith's "harmonic phonology" (Goldsmith, to appear) and George Lakoff's "cognitive phonology" (Lakoff, 1988).

2. One English diagnostic context is the resultative construction (Levin and Rappaport, 1989); under the resultative reading "The door was shut as a result of rolling," the sentence *The door rolled shut* is acceptable, while under the corresponding reading "John was exhausted as a result of working," **John worked exhausted* is unacceptable. This parallels the contrast with transitive verbs between *John wiped the table clean* ("The table was clean as a result of wiping") and **John wiped the table exhausted* ("John was exhausted as a result of wiping"). The argument of intransitive *shut* behaves like the direct object of *wipe* while the argument of *work* behaves like the subject of *wipe*: *shut* is unaccusative while *work* is unergative.

3. Aspect characterizes the internal temporal structure of an event.

4. We take a verb *V* to be telic if "He *V*ed for hours" is unacceptable; we take a verb *V* to be progressive if "He is *V*ing" is acceptable.

5. It is striking that the verbs which are commonly assumed to be prototypical unaccusatives because of their semantic properties actually behave unaccusatively in the fewest numbers of diagnostic contexts. While *exister* "exist" and *être* "be" are strong examples of patient-taking verbs, *exister* behaves unaccusatively in only one context and *être* in only two.

6. These biases are to ill-formedness; a sentence will be unacceptable unless the preferred conditions of the diagnostic context are met, which overcomes the bias. Differences in the strengths of these bias between contexts is one of the factors used to account for mismatches.

7. Because the net inputs to the two hidden units are always the negative of each other, in our implementation we use an equivalent but simpler network in which there is only one hidden unit, which sends the absolute value of its activation to the output unit.

8. The acceptability judgements –, –?, ?, +?, + are taken (rather arbitrarily) to correspond to the numerical values 0.1, 0.3, 0.5, 0.7, 0.9. The same values are used to code the argument and predicate features, except that – is coded as 0 and + is coded as 1.

References

Burzio, L. (1986). *Italian Syntax: A Government-Binding Approach*, Reidel, Dordrecht, Holland.

Chomsky, N. (1965) *Aspects of the theory of syntax*. Cambridge: MIT Press.

Goldsmith, J. (to appear). Harmonic phonology. To appear in J. Goldsmith (ed.) *Proceedings of the Berkeley Workshop on Nonderivational Phonology*. Univ. of Chicago Press.

Lakoff, G. (1988). A suggestion for a linguistics with connectionist foundations. In D. Touretzky, G. Hinton, & T. Sejnowski (eds.), *Proceedings of the 1988 Connectionist Models Summer School*. Morgan Kaufmann Publishers.

Legendre, G. (1989). Unaccusativity in French. *Lingua* **79**, 95–164.

Legendre, G. (forthcoming). *Syntactic vs. semantic accounts of unaccusativity: the need for a new approach.*

Legendre, G., Miyata, Y. & Smolensky, P. (1990) *Harmonic Grammar - A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations.* Technical report #90-5, Institute of Cognitive Science, University of Colorado at Boulder.

Levin, B. & Rappaport, M. (1989). An approach to unaccusative mismatches. *Proceedings of the Nineteenth Meeting of the North Eastern Linguistic Society.*

Perlmutter, D.M. (1978). Impersonal passives and the unaccusativity hypothesis. In *Proceedings of the Fourth Berkeley Linguistic Society Meeting.*

Rosen, C. (1984). The interface between semantic roles and initial grammatical relations. In D.M. Perlmutter and C. Rosen (Eds.), *Studies in Relational Grammar 2*, 38–77. Chicago, IL: University of Chicago Press.

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press/Bradford Books.

Smolensky, P. (1988). On the proper treatment of connectionism. *The Behavioral and Brain Sciences.* **11**, 1–23.

Van Valin, R.D. (to appear). Semantic parameters of split intransitivity. To appear in *Language*.

Zaenen, A. (1989). Unaccusativity in Dutch: An integrated approach. ms, Xerox-PARC and CSLI-Stanford.