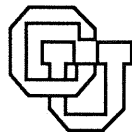


**Can Connectionism Contribute to Syntax?
Harmonic Grammar, with an Application**

**Geraldine Legendre
Yoshiro Miyata
Paul Smolensky**

CU-CS-485-90 July 1990



**University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE**

**ANY OPINIONS, FINDINGS, AND CONCLUSIONS OR RECOMMENDATIONS
EXPRESSED IN THIS PUBLICATION ARE THOSE OF THE AUTHOR(S) AND DO
NOT NECESSARILY REFLECT THE VIEWS OF THE AGENCIES NAMED IN THE
ACKNOWLEDGMENTS SECTION.**

**Can Connectionism Contribute to Syntax?
Harmonic Grammar, with an Application**

Géraldine Legendre^{1,2}

Yoshiro Miyata^{1,3,4}

Paul Smolensky^{1,3,4}

¹Institute of Cognitive Science

²Department of Linguistics

³Department of Computer Science

⁴Optoelectronic Computing Systems Center

University of Colorado at Boulder

Boulder, CO 80309-0430

ICS Technical Report #90-12 CU-CS-485-90 July 1990

To appear in CLS-26I (1990)

"Proceedings of the 26th meeting of the Chicago Linguistics Society"

To appear in CLS-26I (1990)

Can Connectionism Contribute to Syntax? Harmonic Grammar, with an Application

Géraldine Legendre, Yoshiro Miyata, & Paul Smolensky

University of Colorado at Boulder

Introduction

Connectionist networks are computer models based on mathematical idealizations of how information is represented and processed by neural networks in the brain. They consist of large numbers of abstract elements called "units" (idealized neurons) each of which has a numerical "activity level" which it computes on the basis of the activity levels of the other units connected to it in the network. Computer simulations of connectionist networks are usually applied to problems in neuroscience, psychology, or artificial intelligence; in this paper, we report on our use of these networks to address a problem in theoretical linguistics. We will not be discussing neural models of language processing; rather, we will present a new grammar formalism derived from the general mathematical properties of connectionist networks.

We start by outlining the new kind of grammar formalism we have developed, and set out the linguistic problem we will address, namely unaccusativity mismatches in French. This new approach is called *Harmonic Grammar* (HG) for reasons to be explained shortly. The goal of HG 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 the system of soft rules is 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 (1a) with the corresponding soft rule or constraint in (1b).

- (1) 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 .

The difference between the old type of syntactic and lexical rules of French and the new one is illustrated in (2)-(3).

- (2) a. In the *croire* construction, the argument of the embedded verb can never be a deep subject in a well-formed sentence.
- b. In the *croire* construction, if the argument of the embedded verb is a deep subject, then the well-formedness of the sentence is diminished by $C_{CR,Subj}$ ($= 5.0$).
- (3) 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 $C_{fondre,Subj}$ ($= 4.5$).

In Harmonic Grammar, well-formedness is quantitative; it corresponds to the connectionist quantity called *harmony* (Smolensky, 1986; for related harmony-inspired approaches to phonology, see Goldsmith, to appear; Lakoff, 1988; Prince & Smolensky, in preparation). The numerical constants C_X in the soft rules of Harmonic Grammar are computed automatically by presenting the linguistic data to an appropriately designed connectionist network using an appropriate connectionist learning algorithm.¹

The paper is structured as follows: Part I summarizes the problem of characterizing unaccusativity in French, part II presents the HG account, part III describes how the HG account is embodied in a connectionist network, how the numerical constants in the soft rules are computed, and how well our account fares against the data. In part IV, we conclude with a series of general remarks about our approach.

I. The Problem

We assume familiarity with the Unaccusative Hypothesis,² formalized by D. Perlmutter within the framework of Relational Grammar (Perlmutter, 1978) and adopted since by other frameworks, in particular, Government and Binding.

In French, the evidence for the unergative/unaccusative distinction is rich. Legendre (1989) discusses nine diagnostics for unaccusativity, four of which are illustrated below in (4)-(7). These include Object Raising (OR) combined with causative *faire*, *croire* constructions (CR), Participial Absolute constructions (PA), and Reduced Relatives (RR). In these four syntactic contexts, the argument of certain intransitive verbs systematically patterns like the direct object or "2" of a transitive verb while the argument of other intransitive verbs systematically patterns like the subject or "1" of a transitive verb. Legendre (1989) argues that the four constructions illustrated below obey constraints which refer directly to 2-hood (see Legendre, 1989, for details).³

- (4) Object Raising with causative *faire* (OR)
- 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 neige** est facile à faire *fondre*.
The snow is easy to make melt.
 - d. ***Les étudiants** sont faciles à faire *travailler*.
The students are easy to make work.
- (5) *Croire* "believe" constructions (CR)
- a. *Je croyais Marie arrêtée* (par la police).
I believed Marie to be arrested (by the police).
 - b. **Je croyais la police arrêté* Marie.
I believed the police to have arrested Marie.
 - c. *Je croyais Marie déjà sortie*.
I believed Marie to have already gone out.
 - d. **Je croyais Marie éternuée*.
I believed Marie to have sneezed.
- (6) Participial Absolute constructions (PA): (the missing argument in the adverbial clause corresponds to the boldface main clause argument).
- a. **Arrêtée** par la police, Marie a dénoncé ses amis.
Arrested by the police, Marie denounced her friends.
 - b. ***Arrêtée** Marie, la police l'a interrogée.
Arrested Mary, the police interrogated her.
 - c. **Parti** avant l'aube, Pierre est arrivé à destination le jour même.
Gone before dawn, Pierre arrived at his destination on the same day.
 - d. ***Travaillé** toute la nuit, Pierre s'est endormi à 8h du matin.
Worked all night, Pierre fell asleep at 8 a.m.

- (7) Reduced Relatives ("adjectival formation") (RR)
- a. La personne *arrêtée* par la police n'a jamais été relâchée.
The person [who was] *arrested* by the police was never freed.
 - b. *Le policier *arrêté* Ceaucescu a été fêté.
The policeman [who] *arrested* Ceaucescu was celebrated.
 - c. La neige *fondue* a formé de la boue.
The *melted* snow formed mud.
 - d. *Son état *empiré* est alarmant.
His *worsened* condition is alarming.

(4c)-(7c) illustrate unaccusative verbs whose boldface argument behaves like the 2 of transitive verbs (4a)-(7a): they lead to acceptable sentences; by contrast, unergative verbs (4d)-(7d), whose argument behaves like the 1 of transitive verbs (4b)-(7b), lead to unacceptable results. These constructions thus provide diagnostic syntactic contexts in which one class of intransitives is acceptable while another is systematically unacceptable. Considering a large number of intransitive verbs (see Table 4 containing 143 verbs), we see that French exhibits *unaccusative mismatches*: different unaccusativity tests single out distinct but overlapping subsets of unaccusatives. Legendre (1989) argues that no single test can identify the whole class of unaccusatives, but that consideration of all tests can. She formulates a set of necessary syntactic conditions on the various diagnostic contexts of the type exemplified in (4)-(7), leading to the disjunctive necessary and sufficient condition on unaccusativity in (8).

- (8) **Condition on unaccusativity in French** (Legendre, 1989)
An intransitive verb v is unaccusative if and only if v passes at least one unaccusativity test.

Legendre's (1989) analysis illustrates the "syntactic" approach to unaccusativity that has been argued for by Rosen (1984), Perlmutter (1989) and others.

The well-documented existence of mismatches within and across languages has prompted other linguists (including Zaenen, 1989; Van Valin, to appear) to claim that a structurally encoded distinction between unergative and unaccusative verbs is unnecessary: Each diagnostic context is merely sensitive to some semantic or aspectual feature, which cannot be taken as a defining feature for all unaccusative verbs. The "semantic" approach is illustrated in (9) via examples from Italian and Dutch.

- (9) a. According to Van Valin (to appear), auxiliary selection in Italian is sensitive only to the aspectual classification of verbs proposed by Vendler (1967) and elaborated upon by Dowty (1979): state, achievement, and accomplishment verbs select *essere* "be" while activity verbs select *avere* "have" — and that therefore the structural unergative/unaccusative distinction employed by Rosen (1984) and Perlmutter (1989) is unnecessary.
- b. Zaenen (1989) argues that auxiliary selection in Dutch is sensitive only to telicity (or boundedness of the event depicted by the verb) while impersonal passivization in Dutch is sensitive only to volitionality of the argument — and that, again, a non-semantically based structural unergative/unaccusative distinction, as argued for in Perlmutter (1978) is unnecessary.

While the semantic/aspectual approach may be sufficient⁴ to characterize certain Italian and Dutch unaccusativity phenomena that had been previously analyzed in syntactic terms, we suspect that semantic/aspectual distinctions cannot *by themselves* characterize the French phenomena identified as syntactic diagnostics for unaccusativity in (4)-(7). Tables 1 and 2⁵ suggest some of the problems (see Legendre, forthcoming, for a more detailed discussion). Table 1 shows that for each semantic/aspectual property one can find acceptable as well as

unacceptable examples of Object Raising in French; this is true of the other diagnostics as well.

Table 1

Semantic/Aspectual Property	Object Raising:	
	Acceptable	Unacceptable
Accomplishment verb	s'asseoir	aller
Achievement verb	fondre	s'écraser
Activity verb	pleurer	jouer
State verb	s'évanouir	être
Telic verb	s'asseoir	aller
Atelic verb	fondre	boire
Volitional argument	s'asseoir	aller
Non-volitional argument	sécher	exister
Animate argument	s'asseoir	aller
Inanimate argument	sécher	s'écraser

Table 2

Predicate	Argument Features		Predicate Features		Tests			
	Animate	Volitional	Telic	PR	OR	CR	PA	RR
Unaccusative								
fondre	-	-	-	+	+	+	+	+
s'évanouir	+	-	+	+	+	+	+	+
partir	+	+	+	+	?	+	+	+
sortir	+	+	+	?	+?	+	+	+
Unergative								
travailler	+	+	-	+	-	-	-	-
méditer	+	+	-	+	-	-	-	-
éternuer	+	-	-	-	?	-	-	-
empirer	-	-	+	+	-?	-	-	-

Table 2 shows that the value of any of the particular argument or predicate features is neither necessary nor sufficient for the acceptability or unacceptability of these tests; on the other hand, there *are* strong tendencies, which are better illustrated in Table 4. The claim that unaccusativity phenomena are characterizable purely in rather simple semantic/aspectual terms, as has been claimed for Italian and Dutch, is, we believe, untenable for French. On the other hand, while the syntactic analysis provides an account of the data, it, too, is problematic in several respects:

(10) a. *Complexity of the pattern; gradedness of unaccusativity*

Table 4 shows that the pattern is complex. While it is possible to formulate a necessary and sufficient condition for unaccusativity in French, we are still faced with the reality of some verbs passing seven tests, some four, some one, etc. Can we claim that verbs that pass one test are as unaccusative as those that pass seven?

- b. *Discrepancy between semantic/syntactic unaccusativity*
Surprisingly enough (for the adherents of the semantic approach), the verbs which regularly are assumed to be unaccusative because of their semantic characteristics are the least like other unaccusatives in French: *exister* "exist" and *être* "be" are good examples of verbs that select a patient as argument. Of eight syntactic/lexical tests, *être* passes only the "croire" test, though not that straightforwardly; *exister* passes none. One has to appeal to stativity to characterize *exister* as unaccusative.
- c. *Gradability of acceptability judgements*
Precise predictions of the gradability of judgments is a real problem for traditional rule-based approaches. Our data (Table 4) involve five levels of acceptability, +, +?, ?, -?, -: acceptable, marginally acceptable, questionable, marginally unacceptable, unacceptable.
- d. *"Mixed" verbs (Legendre, 1989)*
Certain verbs, including *tousser* "cough", *laugh* "rire", *parler sous la torture* "speak under torture," and a few others, pass one test, OR, with certain kinds of arguments. In a rule-based account, it seems necessary to resort to an ad hoc stipulation, e.g., that the lexicon contains two homophonous verbs, one unergative and the other unaccusative: the unaccusative one is used under OR, the unergative one elsewhere.⁶
- e. *Very strong tendencies*
In spite of these problems, there are very strong tendencies which have to be accounted for. If these tendencies were 100 percent valid, then a rule-based account would work very well. We now proceed with an HG account which formalizes tendencies integrating semantic, aspectual and structural properties.

II. An HG account of unaccusativity in French

The data

Our goal is to provide an account of the acceptability judgements exhibited in Table 4. The data consists of 760 sentence types, a few of which are illustrated in (4)-(7). Each sentence is characterized by a diagnostic test of unaccusativity (a syntactic construction), an embedded intransitive predicate (in bold italics), and the argument of the predicate (in boldface). The grammar assigns a structure, i.e. a deep grammatical relation (GR) to the argument: 1 or 2. Table 3 summarizes all the aspects of a sentence's structural description that are relevant to our analysis.

- (11) a. total of 760 sentence types
b. S = test + **argument** + **predicate** + [structure]
(typeface conventions of (4)-(7))

Table 3

S element	description	network
test	no internal structure, just label: OR, CR, PA, RR	4 input units
argument	two semantic features: \pm AN, \pm VO	2 input units
predicate	two aspectual features: \pm TE, \pm PR [predicate _f] identity: <i>agir</i> , ... [predicate _{id}]	2 input units 143 input units
structure	GR:deep 1 or 2	2 hidden units
acceptability	graded in (0,1); discretized to +, +?, ?, -?, -	1 output unit

Rules

We account for the data with 175 simple rules, outlined in (12), and explained below.

- (12) 175 rules:
- Grammar (32)
 - Structural (8)
 - test / GR (4)
 - a. The *croire* construction strongly prefers the argument of the embedded predicate to be a deep direct object.
 - b. If the test is CR and the GR is 2, add 5.0 to the well-formedness.
 - c. $CR \ \& \ 2 \Rightarrow +5.0$
 - d. $*(CR \ \& \ 1)$
 - e. $[CR \ \& \ 1 \Rightarrow -5.0 \quad (2b)]$
 - argument / GR (2)
 - f. Animate arguments strongly prefer to be deep subjects.
 - g. If the argument is +AN and the GR is 2, add -4.4 to the well-formedness.
 - h. $+AN \ \& \ 2 \Rightarrow -4.4$
 - i. $*(+AN \ \& \ 2)$
 - j. $[-AN \ \& \ 2 \Rightarrow 0]$
 - predicate_f / GR (2)
 - k. Telic predicates prefer their arguments to be deep direct objects.
 - l. $*(+TE \ \& \ 1)$
 - m. $+TE \ \& \ 2 \Rightarrow +2.7$
 - Nonstructural (24)
 - test / argument (8)
 - n. $CR \ \& \ +AN \Rightarrow +5.1$
 - test / predicate_f (8)
 - o. $CR \ \& \ +TE \Rightarrow +1.2$
 - argument / predicate_f (4)
 - p. $+AN \ \& \ +TE \Rightarrow +6.8$
 - test (4)
 - q. $CR \Rightarrow -10.6$
 - Lexicon (143)
 - Structural (143)
 - predicate_{id} / GR (143)
 - r. *fondre* strongly prefers its argument to be deep direct object.
 - s. $*(fondre \ \& \ 1)$
 - t. $fondre \ \& \ 2 \Rightarrow +4.5 \quad (3b) \quad \text{— unaccusative bias}$
 - u. *éternuer* strongly prefers its argument to be a deep subject.
 - v. $*(éternuer \ \& \ 2)$
 - w. $éternuer \ \& \ 2 \Rightarrow -3.4 \quad \text{— unergative bias}$

There are 32 grammatical rules and 143 lexical rules. The grammatical rules consist of eight structural and 24 non-structural rules. The structural rules refer to the GR borne by the argument of the embedded predicate: four of these rules refer to the test, two to the semantic

features of the argument (ANimacy, VOLitionality), and two to the aspectual features of the predicate (TElicity, PRogressivity).⁷

One structural rule referring to the test is given in four notations in (12a-d). In words, the rule is stated in (12a). The vague phrase "strongly prefers" is made precise in (12b). (On our arbitrarily chosen scale of well-formedness, the smallest difference in well-formedness between completely acceptable and completely unacceptable sentences is about 3 units.) (12b) is expressed more compactly in (12c), which is the form in which we will express the other rules. This rule can also be regarded as a soft version of the well-formedness filter in (12d); the corresponding rule appearing in our HG account, (12e), could also be viewed as a quantified markedness condition. (12e) was expressed verbally in (2b). In our account, to each structural rule referring to 2-hood, like (12c), there is a "mirror rule," like (12e), that refers to 1-hood, containing a numerical constant that is the negative of the constant appearing in the corresponding 2-hood rule. (In our counting of rules, we have not included mirror rules, since they can be viewed as generated automatically via a meta-rule.) The other three structural rules referring to tests involve replacing CR with OR, PA, and RR, each rule having its own numerical parameter. (The differences across tests of these and other test-dependent parameters are what makes unaccusative mismatches possible.)

The other structural rules that are central to the syntactic analysis of unaccusativity correspond in our account to the 143 lexical rules stating the preferences of individual lexical items for the deep GR of their argument. An example is illustrated in three different notations in (12r-t). This lexical rule, stated in English in (12r), a soft version of the filter in (12s), is stated more precisely using our standard notation in (12t) (see also (3b)). Note that this rule says that, independent of its aspectual features, *fondre* "to melt" strongly prefers a 2; in our account, *fondre* is structurally unaccusative. By contrast, *éternuer* "to sneeze" is structurally unergative, as expressed variously in (12u-w). There is one such rule for each of the 143 predicates we consider.

The remaining types of rules are illustrated in (12f-q). An example of a structural rule referring to the semantic features of the argument is given in (12f-i), using four notations, parallel with (12a-d). As indicated in (12j), all our rules referring to features apply only to the + value; features with value – do not contribute to the computation of well-formedness.⁸

Rule interaction

The rules are combined to predict the acceptability of a given sentence as follows (see (13)).

- (13) Computation of well-formedness or harmony:
- a. $H = H_{nonstructural} + H_{structural}$
 - b. $H_{nonstructural}$ = sum of contributions from all applicable nonstructural rules in grammar
 - c. $H_{structural}$ = maximum of H_1 and H_2
 $H_{1,2}$ = sum of contributions from all applicable structural rules in grammar and lexicon, if assign GR 1,2
 - d. **Harmonic structural assignment:** the grammar and lexicon assign to an input S the structural description that maximizes total harmony, H .
 - e. acceptability is computed from H : $H < 0 \Rightarrow -$; $H > 0 \Rightarrow +$; $H = 0 \Rightarrow ?$

A sentence's well-formedness, measured by the connectionist quantity H , *harmony*, consists of two parts, the *nonstructural harmony* $H_{nonstructural}$, and the *structural harmony* $H_{structural}$, which are added together numerically (13a). Each applicable nonstructural rule contributes a certain numerical value; the sum of these numbers is the nonstructural harmony, which can be positive or negative. The structural harmony is a bit more complex, because the structural rules refer to

the GR of the argument, which is of course not given by the input sentence directly: it must be assigned by the grammar, as follows (see 13c). First, hypothesize that the GR to be assigned is 2; then add up all the contributions of the applicable structural rules, getting a number H_2 . If H_2 is positive, then the grammar assigns GR 2 to the argument, and the structural harmony is H_2 . Otherwise, the grammar assigns GR 1 to the argument; if we now add up all the applicable structural rules — the mirror rules like (12e) referring to GR 1 — then we get the structural harmony H_1 , which is just the negative of H_2 . This means of assigning deep GRs is a simple special case of the most fundamental principle of Harmonic Grammar, expressed in (13d). Finally, the acceptability of a sentence is computed from the total harmony H : qualitatively, this is described in (13e); the quantitative relation is a detail of the connectionist implementation, discussed next in part III.⁹

The general principle (13d) enables highly context-dependent structural assignments. In our HG account of unaccusativity, for example, the deep GR assigned to the argument of the embedded predicate is sensitive to the syntactic construction in which it is embedded and the semantic features of the argument, as well as the aspectual features of the predicate and the identity of the predicate; the latter contributes a preference or *bias* for the deep GR assigned to its argument, but this is merely one of several factors all of which determine the harmony-maximizing choice of deep GR.

III. The connectionist implementation

The account of French unaccusativity described in the preceding section is computationally implemented as a computer-simulated connectionist network. This allows automatic computation of the predicted well-formedness of sentences, and, most importantly, allows automatic determination from the data of the numerical parameters appearing in the soft rules.

The connectionist network, shown in Figure 1, contains the units described in the last column of Table 3. A sentence is input to the network by appropriately setting the activities of the input units: the test or "context" unit corresponding to the appropriate syntactic construction is given activity 1, the other three test units are given activity 0; among the 143 predicate-identity units, the one corresponding to the appropriate embedded predicate is given activity 1 and the others activity 0; for the two argument and two predicate feature units, the unit for each feature is given the numerical activity value encoding that feature's value: 1 for +, 0 for -, and .3, .5, and .7 for -, ?, and +?, respectively.

In addition to the input units, there is a single output unit whose value ranges between 0 (most unacceptable) and 1 (most acceptable). There are also two interior or "hidden" units, labelled **1** and **2**: if the network decides to activate unit **1** (or **2**), that implements the grammar's assignment of deep 1 (or 2) to the argument of the embedded intransitive. How this decision is made will be discussed momentarily. The final set of units denote conjunctions of input units; in Figure 1, for example, the unit for the conjunction of AN and TE is explicitly labelled **AN&TE**. The activity value of **AN&TE** is the numerical product of the activations of the **AN** and **TE** units; if both AN and TE are +, the corresponding units have activity 1 and so does **AN&TE**; if either AN and TE are -, the activity of **AN&TE** will be 0.

The 175 grammatical and lexical rules in our account are each directly encoded in the network as a connection, whose weight or strength is the numerical parameter in the corresponding rule. The rule (12p) "+AN & +TE \Rightarrow +6.8" is implemented by a connection of strength +6.8 from the **AN&TE** conjunction unit to the output (acceptability) unit. The rule "CR & 2 \Rightarrow +5.0" is implemented by a connection of strength +5.0 from the **CR** unit to the **2** unit.

With these connections, the activation flowing into the **2** unit (the sum over all input units of their activation times the strength of the connection to **2**) is exactly the harmony H_2 that would result from the structural rules in the grammar and lexicon if the argument were assigned deep GR 2. Similarly, the activation flowing into the **1** unit is H_1 , the negative of H_2 . Whichever hidden unit receives more activation will attain an activity value equal to the activation it receives, and the other hidden unit will have its activation set to 0; the two hidden units form what connectionists call a "winner-take-all" group. If hidden unit **2** (or **1**) wins this competition, then the argument is assigned the deep GR 2 (or 1). The activity value of the winning hidden unit, which is just the structural harmony $H_{structural}$, is then sent up to the output unit (along a connection of strength +1).

The output unit receives activation from the winning hidden unit and also directly from the input and conjunction units. The former activation is $H_{structural}$, and the total of the latter activation is the non-structural harmony $H_{non-structural}$; following (13a), these two are added together to form the total activation flowing into the output unit: this is the total harmony H or well-formedness of the sentence, according to the soft rules embodied in the connections. This output unit achieves a value between 0 and 1 according to a rather arbitrarily chosen function (the logistic) that is typically used in connectionist networks: $acceptability = 1/(1+e^{-H})$. Following (13e), if H is strongly negative, the output is close to 0 (most unacceptable); if H is strongly positive, the output is close to 1 (most acceptable); to match the predicted acceptability against the informants' judgements, we categorize the output as -, -?, ?, +?, and +, whenever the value is closest to .1, .3, .5, .7, and .9, respectively.

The particular values of the 175 independent numerical parameters that appear in our soft rules are determined automatically from the informants' judgements by training the connectionist network, which has an exactly corresponding 175 independent connection strengths. For each of our 760 sentences, we input the sentence by assigning the appropriate activity values to the input units, and we tell the network what the correct output value is, according to the informants. The network gradually adjusts the connection strengths, attempting to make its predictions match the data as closely as possible. The particular learning algorithm used is a version of "back-propagation" (Rumelhart, Hinton, & Williams, 1986) that appropriately handles our particular (and somewhat unusual) kind of hidden units.

The best result of learning to date is a set of numerical parameters for soft rules (including the examples in this paper) that correctly accounts for the sign of the acceptability of all but two of the 760 sentence types. That is, when the informants' rating is + or +?, so is the network's, and similarly for - and -?, except for two cases, each involving marginal judgements.¹⁰

IV. Concluding remarks

HG and the problems of (10)

The HG formalism allows us to address, at least in part, the problems identified in (10):

- (14) a. *Complexity of the pattern; gradedness of unaccusativity*
The HG computational framework allows precise predictions involving a very complex interaction of numerous rules, in which the structural unaccusativity of a predicate is a numerical bias on deep GR assignment.
- b. *Discrepancy between semantic/syntactic unaccusativity*
This is what we would expect in a rule system such as that proposed here in which semantic and syntactic rules interact heavily.
- c. *Gradability of acceptability judgements*
This is of course handled completely naturally in a framework of numerical well-formedness computation.

d. *"Mixed" verbs*

The full story here is rather complex, but one aspect is transparent: since our account involves context-sensitive GR assignment, it will happen that with some predicates, whether the argument is assigned deep grammatical relation 1 or 2 will change depending on the embedding context, e.g., the syntactic test.

e. *Very strong tendencies*

Our grammar and lexicon are exactly explicit sets of rules quantifying these tendencies (in a subtle statistical fashion), allowing them to be used to make precise predictions.

Further research on unaccusativity

To indicate what further investigations can be pursued within the Harmonic Grammar framework, we briefly sketch ongoing and planned research on unaccusativity in French. We are currently analyzing the network's solution to a number of problematic examples in the data, to gain further insight into the explanations provided by this account. We are testing the importance of various semantic/aspectual features and their interactions through rules by building competing accounts using different features and allowing different interactions. We plan to test various aspects of the generality of our account, e.g., testing the grammar's extensibility to new predicates to see whether, without changing the grammatical rules, we can successfully add new verbs to the lexicon, and to see indeed if this can be achieved by training the network on a subset of our current 143 predicates, and then merely extending the lexicon to cover the remaining verbs. We intend to test purely semantic accounts with richer semantic representations involving more features, to see whether such an approach can exploit the power of soft rules to account for these data without postulating nonsemantically determined structural preferences for individual intransitive verbs, and possibly without even postulating a structural GR 1 / 2 distinction at all. We plan to extend the account to transitive verbs, to examine in the new framework the parallelism on which the Unaccusative Hypothesis is based, between the 1s and 2s of transitives and the arguments of unergative and unaccusative intransitives.

Relation to other approaches to unaccusativity integrating syntax and semantics

The degree of syntactic/semantic interaction exhibited in this particular Harmonic Grammar treatment of unaccusativity in French is considerably stronger than in the other approaches to unaccusativity of which we are aware. In purely semantic approaches, the interaction is of course nil, since well-formedness is determined directly from semantic properties and constraints. In the syntactic approach, well-formedness is determined by syntactic conditions, explicit and typically simple (e.g., "In OR the raisee must be a 2 in all strata prior to raising"; Legendre, 1989), with, often, inexplicit semantic or aspectual conditions presumed to operate as well (e.g., "In OR, the raisee must be sufficiently generic"). In Zaenen's (1989) integrated approach, semantic features determine syntactic features which then in turn directly determine well-formedness. In the HG account described here, however, neither syntactic nor semantic properties of the input alone can determine acceptability, and the interaction is not a simple conjunction of separate syntactic and semantic conditions: the well-formedness is directly contributed to by both structural and non-structural considerations, some with a semantic basis and others involving non-semantically governed structural preferences of individual lexical items.

Mathematical derivation of the HG formalism from connectionist principles

The rule formalism of Harmonic Grammar can be mathematically derived from basic connectionist principles. The full derivation can be found in Legendre, Miyata, & Smolensky (1990b); here, we simply give a sketch.

Following Smolensky (1988), the fundamental assumption on cognitive architecture is that symbolic rule systems are higher-level, in various ways approximate, descriptions of the global behavior of lower-level connectionist networks. We assume that the lower level network is characterized, at least approximately, by certain basic principles, given in (15a-b). In addition, we make two assumptions about the relation between this lower-level model and theoretical linguistics, (15c-d). Beyond these assumptions, we do not further specify the lower-level network; in particular, unlike in most connectionist research, we don't specify the lower-level network to the great detail required to simulate it on a computer.

- (15) a. Structured data (such as the structural descriptions of sentences) is represented as distributed patterns of activity using a general technique called "tensor product representations" (Smolensky, in press).
- b. The activation spread that constitutes processing in the network achieves the end result of maximizing harmony (or minimizing "energy"; see, e.g., Cohen & Grossberg, 1983; Golden, 1986, 1988; Hinton & Sejnowski, 1983; Hopfield, 1982; Smolensky, 1983, 1986)
- c. The connectionist harmony of the distributed representation of the structural description of a sentence can be taken as a measure of its linguistic well-formedness.
- d. The overall structure of the tensor product representation and the overall structure of the constraints among constituents that is embodied in the harmony function correspond to structural descriptions and principles of theoretical linguistics.

From the assumptions in (15), it is possible to mathematically derive a formal higher-level description of the representation and processing in the lower-level network; this higher-level description, as it happens, is embodied in another connectionist network, the one illustrated in Figure 1. There are many differences between the lower- and higher-level networks: the former is partially specified and not simulated, while the latter is fully specified and simulated; the former uses distributed representations while the latter uses local representations; in the former, harmony is distributed throughout the net, while in the latter there is a single "output" unit that computes harmony (acceptability); the latter is transparently interpreted as an implementation of a Harmonic Grammar, while the former is not. Despite all these differences, the two networks are *isoharmonic*: they compute the same harmony function, and that is what is relevant for assessing well-formedness.¹¹

It is worth emphasizing that Harmonic Grammar is a general technique for taking hypotheses about linguistic representations and interactions and automatically generating accounts of data from them. For example, as mentioned above, it is straightforward to generate other HG accounts of unaccusativity in French by adopting different basic hypotheses: e.g., that there is no idiosyncratic structural preference for individual predicates in the lexicon, or that there is no element of the structural description corresponding to deep 1 or 2, or that there are no grammatical constraints between the aspectual features of a predicate and the animacy or volitionality of its argument. None of the substantive linguistic assumptions that underly the particular account of unaccusativity in French presented here are in any way commitments of Harmonic Grammar; on the contrary, at this point, the commitments of the approach are formal ones, such as the use of soft rules in a computation of harmony, and principle (13d) for harmonic

structural description assignment — and to the more fundamental principles from which these are derived: the assumptions (15) concerning the underlying connectionist substrate and its relation to theoretical linguistics.

Applicability to other areas of linguistics

It should be pointed out that Harmonic Grammar is potentially applicable to linguistic well-formedness conditions generally, and not just to the study of syntactic or semantic well-formedness. In phonology, for example, the approach can be used to improve the explanatory adequacy of the theory (Prince & Smolensky, in preparation), making the further point that Harmonic Grammar offers opportunities for improving the explanatory and not just the descriptive adequacy of linguistic theory.

Through its roots in connectionism, a computational framework, Harmonic Grammar provides a strong link between descriptive and computational linguistics. The HG measure of well-formedness, harmony, is integral to *computing with the grammar*, since connectionist processing *is just* an algorithm for maximizing harmony.

Connectionism also provides Harmonic Grammar with opportunities for studies of language acquisition, for, while the learning algorithm used in this paper is not plausible as a model of human language acquisition (e.g., negative evidence is crucial), a variety of other connectionist learning algorithms are available which may allow further development of language learning models.

Finally, we believe Harmonic Grammar can provide a much-needed link between theoretical linguistics and connectionist cognitive models of language processing.

Summary

We have presented Harmonic Grammar, a formalism in which grammatical and lexical constraints are embodied in soft rules, which quantify the consequences for well-formedness of satisfying or violating what may be regarded as markedness conditions or preferences. This formalism is mathematically derived from a few fundamental principles concerning connectionist representation and processing, and their relation to theoretical linguistics. Harmonic Grammar provides a general framework within which to express specific hypotheses about structural descriptions and constraints for a specific linguistic domain, implement them in a connectionist network, train this network on acceptability judgements, and interpret the result as a collection of soft rules. We have illustrated the approach with an account of unaccusativity mismatches in French that provides a striking degree of coverage of problematic data, using strongly interacting syntactic and semantic rules.

By working within a formalism that is mathematically derived from connectionism, a cognitively (and statistically) based computational framework, this approach to grammar is in a good position from which to pursue integration with other cognitive language models (e.g., acquisition, real-time processing, and neurolinguistic models). However, unlike most connectionist research on language, Harmonic Grammar incorporates a number of attractive features of formal symbolic linguistics; it allows us to:

- 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 theoretical linguistics.

Acknowledgements

GL acknowledges the support of a Junior Faculty Development Award from the Council on Research and Creative Work, University of Colorado, Boulder. YM and PS have 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. PS also acknowledges the support of NSF grant IRI-8609599 and a grant from the Sloan Foundation's computational neuroscience program.

We are grateful for comments and hospitality from Jeff Elman and the UCSD Center for Research in Language, feedback from our colleagues in the University of Colorado Institute of Cognitive Science, and for helpful discussions with Tom Bever, Gilles Fauconnier, John Goldsmith, Mary Hare, Judy Kegl, Jim Martin, Mike Mozer, David Perlmutter, Steven Pinker, Alan Prince, Susan Steele, and David Touretzky.

The authors are listed in alphabetical order. Requests for reprints can be sent to GL, Department of Linguistics, Campus Box 295, or YM and PS, Department of Computer Science, Campus Box 430, all at the University of Colorado, Boulder, CO 80309.

Footnotes

1. 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.

2. The Unaccusative Hypothesis claims that intransitive verbs are of two kinds: **unergative** verbs, which take an initial 1 or deep subject, and **unaccusative** verbs, which take an initial 2 or deep direct object.

3. The remaining five diagnostics for unaccusativity discussed in Legendre (1989), including parallel transitive structures, cliticization of the embedded 3 ("indirect object") in causative constructions, auxiliary selection, nominalizations, and stativity, do not pertain to 2-hood per se (except for the diagnostic based on parallel transitive structures) and apply only to small subsets of unaccusative verbs. In this first HG account, we have limited ourselves to the four most productive syntactic diagnostics that are sensitive to 2-hood.

4. The use of *may* is intentional here. We note that most analyzes of unaccusativity phenomena rely on samples of verbs too small to draw the far-reaching consequences that have customarily been drawn.

5. The verbs exemplified in Tables 1 and 2 are translated as follows: *s'asseoir* "to sit down", *aller* "to go", *fondre* "to melt", *s'écraser* "to crash", *pleurer* "to cry", *jouer* "to play", *s'évanouir* "to faint", *être* "to be", *boire* "to drink", *sécher* "to dry off", *exister* "to exist", *partir* "to leave", *sortir* "to go out", *travailler* "to work", *méditer* "to meditate", *éternuer* "to sneeze", *empirer* "to worsen".

6. One might object that the problem is with the analysis of OR. Legendre (1989) discusses an alternative account of OR and concludes that a purely semantic account of OR does not work in French.

7. These features have been identified in the literature as relevant to these phenomena.

8. This is merely a formal convenience; notational variants of this account using various treatments of – and + could be readily constructed.

9. Since the structural harmony is always positive, each test provides an overall negative bias on the well-formedness, via rules such as (12q). If the structural and non-structural well-formedness is adequate to overcome this negative bias, the sentence will be acceptable, otherwise not. One source of unaccusativity mismatches is the differences across tests of the strength of this negative bias.

10. To achieve this, we started the connections at weights that correspond roughly to a "purely syntactic" account of unaccusativity; prior to learning, all weights are set to zero, except that the tests all prefer GR 2, and individual predicates prefer GR 2 if they pass most of the tests, and deep 1 if they fail most of the tests. This seems to start the weights off in the right ballpark, even though prior to learning, the sign of acceptability is wrongly predicted for 92 sentences. After learning, the account is, of course, very different. For further discussion of the connectionist implementation, see Legendre, Miyata, & Smolensky (1990a).

11. What is the real contribution of the lower-level network, if all our predictions of well-formedness come from the higher-level network? The lower-level network *explains* many of the features of the higher-level network that would otherwise be totally ad hoc: the particular use of conjunction units (why two-way conjunctions and not three-way conjunctions?; why don't the conjunctive units feed into the hidden units?), the particular (unusual) kind of hidden units (why linear? why winner-take all?), the explicit unit for acceptability, the use of local representations. All these are mathematical consequences of assumptions about the lower-level network.

References

- Cohen, M. A. & Grossberg, S. (1983). *Absolute stability of global pattern formation and parallel memory storage by competitive neural networks*. IEEE Trans SMC-13, 815–825.
- Golden, R. M. (1986) The "Brain-State-in-a-Box" Neural Model Is a Gradient Descent Algorithm. *Mathematical Psychology*, 30-1, 73–80.
- Golden, R. M. (1988) A Unified Framework for Connectionist Systems. *Biological Cybernetics*, 59, 109–120.
- Goldsmith, J. (to appear). Harmonic phonology. To appear in J. Goldsmith (ed.) *Proceedings of the Berkeley Workshop on Nonderivational Phonology*. Univ. of Chicago Press.
- Hinton, G.E. & Sejnowski, T.J. (1983). Analyzing cooperative computation. *Proceedings of the Fifth Annual Conference of the Cognitive Science Society*. Rochester, NY.
- Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA* 79, 2554–2558.
- 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. (1990a). Harmonic Grammar — A formal multi-level connectionist theory of linguistic well-formedness: An application. Technical Report 90-4, Institute of Cognitive Science, University of Colorado at Boulder.
- Legendre, G., Miyata, Y. & Smolensky, P. (1990b) 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.
- Perlmutter, D.M. (1978). Impersonal passives and the unaccusativity hypothesis. *Proceedings of the Fourth Berkeley Linguistic Society Meeting*.
- Perlmutter, D. M. (1989). Multiattachment and the unaccusative hypothesis: The perfect auxiliary in Italian. *Probus: Journal of Latin and Romance Linguistics* 1, 63–119.

- Prince, A. & Smolensky, P. (in preparation). Universal phonology and harmony theory.
- 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. (1983). Schema selection and stochastic inference in modular environments. *Proceedings of the National Conference on Artificial Intelligence*. Washington, DC.
- Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. 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.
- Smolensky, P. (in press). Tensor product variable binding and the representation of symbolic structures in connectionist networks. *Artificial Intelligence*.
- 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.

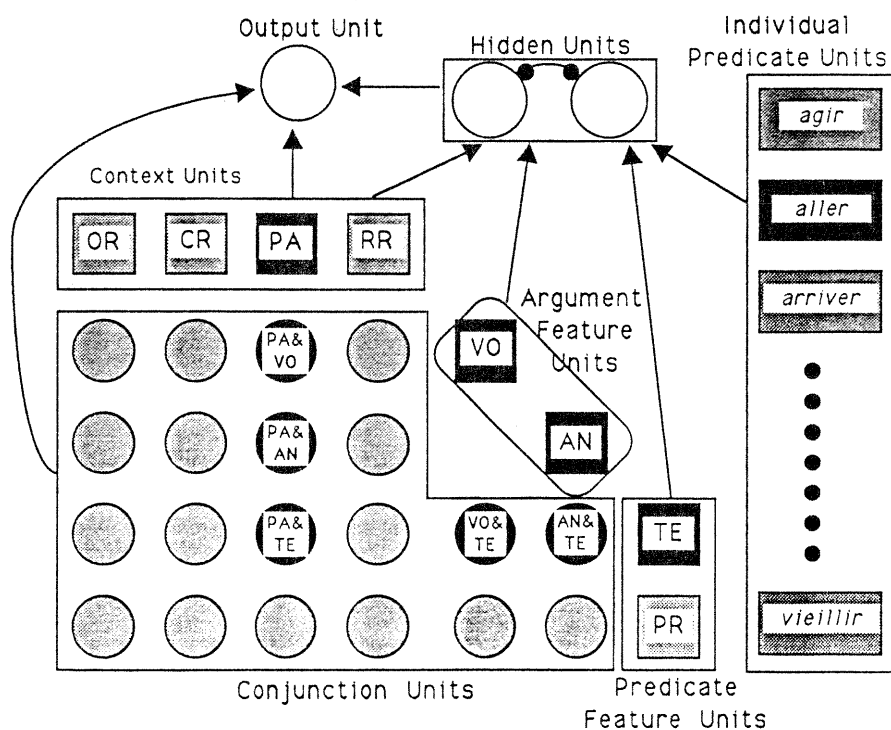


Figure 1

Table 4

OR CR PA RR Predicate	OR CR PA RR Predicate	OR CR PA RR Predicate
- - - - - + sévir	- + + + + + - ? se réfugier	+ - - - + + + réagir
- - - - - + résister	- + + + + + - surgir	+ - - - + + + céder
- - - - - + persister	- ? - - - - - ? subsister	+ + ? + ? + ? - - - + se modifier
- - - - - + continuer	- ? - - - - - + empirer	+ + ? + ? + ? - - - + se dissiper
- - - - - + ouvrir sur	- ? - - - - - + + + s'ouvrir	+ + ? + + ? - - - + peler
- - - - - + exister	- ? - - - - - + + - ronchonner	+ + ? + + - - + vieillir
- - - - - + souffrir	- ? - - - - - + + + exploser	+ + - + ? - - - + pâlir
- - - - - + persévérer	- ? - - - - - + + + changer d'avis	+ + + + ? + - - - + se disperser
- - - - - + marcher	- ? - - - - - + + + s'esquiver	+ + + ? + ? - + + + changer
- - - - - + survivre	- ? - ? ? ? + + + + s'enfuir	+ + + ? + - - - + pousser
- - - - - + succéder	- ? + ? - + ? - - - ? peler	+ + + ? + - - - + éclater
- - - - - + subsister	- ? + ? + + + + venir	+ + + + + - - - + jaillir
- - - - - + régner	- ? + - + ? ? + + + pâlir	+ + + + + - - - + exploser
- - - - - + persister	- ? + - + - - - + grandir	+ + + + + - - - + tomber
- - - - - + mendier	- ? + + + - - - + se retirer	+ + + + + - - - + sombrer
- - - - - + être	- ? + + + - - - + s'ouvrir	+ + + + + - - - + sécher
- - - - - + (dés)obéir	- ? + + + - - - + geler	+ + + + + - - - + s'agrandir
- - - - - + cesser	- ? + + + - - - + parvenir	+ + + + + - - - + rouiller
- - - - - + - ? déambuler	- ? + + + - - - + se retirer	+ + + + + - - - + rôtir
- - - - - + - ? exagérer	? - - - - - + + - - - éternuer	+ + + + + - - - + ralentir
- - - - - + + + travailler	? - - - - - + + + + reculer	+ + + + + - - - + paraître
- - - - - + + + téléphoner	? + + + - - - + vieillir	+ + + + + - - - + noircir
- - - - - + + + souffler	? + + + - - - + brunir	+ + + + + - - - + monter
- - - - - + + + sévir	? + + + - - - + naître	+ + + + + - - - + mariner
- - - - - + + + sautiller	? + + + - - - + guérir	+ + + + + - - - + geler
- - - - - + + + rouspéter	? + + + - - - + monter	+ + + + + - - - + fondre
- - - - - + + + r'éver	? + + + - - - + descendre	+ + + + + - - - + fermer
- - - - - + + + résister	? + + + - - - + entrer	+ + + + + - - - + disparaître
- - - - - + + + répondre	? + + + - - - + arriver	+ + + + + - - - + descendre
- - - - - + + + réfléchir	? + + + - - - + apparaître sur scène	+ + + + + - - - + cuire
- - - - - + + + ralentir	? + + + - - - + partir	+ + + + + - - - + cristalliser
- - - - - + + + penser*	+ ? - - - - - + + + durer	+ + + + + - - - + brunir
- - - - - + + + parler	+ ? - - - - - + + + ? capituler	+ + + + + - - - + brûler
- - - - - + + + mentir	+ ? - - - - - + + + ? rougir	+ + + + + - - - + bouillir
- - - - - + + + méditer	+ ? - - - - - + + + ? parler sous la torture	+ + + + + - - - + baisser
- - - - - + + + jouer	+ ? - - - - - + + + - - - trébucher	+ + + + + - - - + augmenter
- - - - - + + + jongler	+ ? + ? + + ? - - - + s'éloigner	+ + + + + - - - + entrer
- - - - - + + + insister	+ ? + ? + + ? + + + + s'éloigner	+ + + + + - - - + apparaître
- - - - - + + + hésiter	+ ? + + ? + - - - + surgir	+ + + + + - - - + sauter
- - - - - + + + gémir	+ ? + + + - - - + sortir	+ + + + + - - - + ? arriver
- - - - - + + + errer	+ ? + + + - - - + venir	+ + + + + - - - + s'éteindre
- - - - - + + + danser	+ ? + + + - - - + s'effacer	+ + + + + - - - + s'envoler
- - - - - + + + continuer	+ ? + + + - - - + tomber	+ + + + + - - - + se casser
- - - - - + + + cogner	+ ? + + + - - - + disparaître	+ + + + + - - - + se briser
- - - - - + + + circuler	+ ? + + + - - - + se dissimuler	+ + + + + - - - + parvenir
- - - - - + + + agir	+ ? + + + - - - + sombrer	+ + + + + - - - + partir
- - - - - + + + sauter	+ ? + + + - - - + ? se blottir	+ + + + + - - - + naître
- - - - - + + + renoncer	+ ? + + + - - - + ? se recroqueviller	+ + + + + - - - + guérir
- - - - - + + + trotter	+ ? + + + - - - + ? sortir	+ + + + + - - - + éclore
- - - - - + + + durer	+ ? + + + - - - + s'exiler	+ + + + + - - - + cicatriser
- - - - - + + + boiter	+ - - - - - + + + souffler	+ + + + + - - - + changer
- - - - - + + + bégayer	+ - - - - - + + + régner	+ + + + + - - - + casser
- - - - - + + + baisser	+ - - - - - + + + reculer	+ + + + + + + + se réunir
- - - - - + + + balbutier	+ - - - - - + + + courir	+ + + + + + + + se noyer
- - - - - + + + flâner	+ - - - - - + + + couler	+ + + + + - - - + grossir
- - - - - + + + courir	+ - - - - - + + + circuler	+ + + + + - - - + s'évanouir
- - - - - + + + aller	+ - - - - - + + + cesser	+ + + + + - - - + mourir
- - - - - + + + - ? courir	+ - - - - - + + + tousser	+ + + + + + + + se décourager
- - - - - + + + rester	+ - - - - - + + + régresser	+ + + + + + + + se cacher
- ? + ? ? + + - - - tomber	+ - - - - - + + + pleurer	+ + + + + + + + s'évader
- + + + + - - - - - décevoir	+ - - - - - + + + marcher	+ + + + + + + + se disperser
- + + + + - - - - - périr	+ - - - - - + + + se taire	+ + + + + + + + s'asseoir
- + + + + + + - - - s'effacer	+ - - - - - + + + se saouler	
- + + + + + + - - - rester	+ - - - - - + + + rire	