COLING 2010

23rd International Conference on Computational Linguistics

Proceedings of the

8th Workshop on Asian Language Resources

21-22 August 2010
Beijing, China

Produced by
Chinese Information Processing Society of China

To order the CD of Coling 2010 and its Workshop Proceedings, please contact:

Chinese Information Processing Society of China
No.4, Southern Fourth Street
Haidian District, Beijing, 100190
China
Tel: +86-010-62562916
Fax: +86-010-62562916
cips@iscas.ac.cn

Unaccusative/Unergative Distinction in Turkish: A Connectionist Approach

Cengiz Acartürk
Middle East Technical University
Ankara, Turkey
acarturk@acm.org

Deniz Zeyrek
Middle East Technical University
Ankara, Turkey
dezeyrek@metu.edu.tr

Abstract

This study presents a novel computational approach to the analysis of unaccusative/unergative distinction in Turkish by employing feed-forward artificial neural networks with a backpropagation algorithm. The findings of the study reveal correspondences between semantic notions and syntactic manifestations of unaccusative/unergative distinction in this language, thus presenting a computational analysis of the distinction at the syntax/semantics interface. The approach is applicable to other languages, particularly the ones which lack an explicit diagnostic such as auxiliary selection but has a number of diagnostics instead.

1 Introduction

Ever since Unaccusativity Hypothesis (UH, Perlmutter, 1978), it is widely recognized that there are two heterogeneous subclasses of intransitive verbs, namely unaccusatives and unergatives. The phenomenon of unaccusative/unergative distinction is wide-ranging and labeled in a variety of ways, including active, split *S*, and split intransitivity (SI). (cf. Mithun, 1991).[1]

Studies dealing with SI are numerous and recently, works taking auxiliary selection as the basis of this syntactic phenomenon have increased (cf. McFadden, 2007 and the references therein). However, SI in languages that lack explicit syntactic manifestations such as auxiliary selection has been less studied.[2] Computational approaches are even scarcer. The major goal of this study is to discuss the linguistic issues surrounding SI in Turkish and present a novel computational approach that decides which verbs are unaccusative and which verbs are unergative in this language. The computational approach may in turn be used to study the split in lesser-known languages, especially the ones lacking a clear diagnostic. It may also be used with well-known languages where the split is observed as a means to confirm earlier predictions made about SI.

2 Approaches to Split Intransitivity (SI)

Broadly speaking, approaches to the SI may be syntactic or semantic. Syntactic approaches divide intransitive verbs into two syntactically distinct classes. According to the seminal work of Perlmutter (1978), unaccusative and unergative verbs form two syntactically distinct classes of intransitive verbs. Within the context of Relational Grammar, Perlmutter (1978) proposed that unaccusative verbs have an underlying object promoted to the subject position, while unergative verbs have a base-generated subject. This hypothesis, known as the Unaccusativity Hypothesis (UH) maintains that the mapping of the sole argument of an intransitive verb onto syntax as subject or direct object is semantically predictable. The UH distinguishes active or activity clauses (i.e., unergative clauses) from unaccusative ones. Unergative clauses include

[1] In this paper, the terms *unaccusative/unergative distinction* and *split intransitivity (SI)* are used interchangeably.

[2] An exception is Japanese. For example see Kishimoto (1996), Hirakawa (1999), Oshita (1997), Sorace and Shomura (2001), and the references therein. Also see Richa (2008) for Hindi.

Proceedings of the 8th Workshop on Asian Language Resources, pages 111–119,
Beijing, China, 21-22 August 2010.

willed or volitional acts (*work*, *speak*) and certain involuntary bodily process predicates (*cough*, *sleep*); unaccusative clauses include predicates whose initial term is semantically patient (*fall*, *die*), predicates of existing and happening (*happen*, *vanish*), nonvoluntary emission predicates (*smell*, *shine*), aspectual predicates (*begin*, *cease*), and duratives (*remain*, *survive*).

From a Government and Binding perspective, Burzio (1986) differentiates between two intransitive classes by the verbs' theta-marking properties. In unaccusative verbs (labeled 'ergatives'), the sole argument is the same as the deep structure object; in unergative verbs, the sole argument is the same as the agent at the surface. The configuration of the two intransitive verb types may be represented simply as follows:

Unergatives:	NP [$_{VP}$ V]	John ran.
Unaccusatives:	[$_{VP}$ V NP]	John fell.

In its original formulation, the UH claimed that the determination of verbs as unaccusative or unergative somehow correlated with their semantics and since then, there has been so much theoretical discussion about how strong this connection is. It has also been noted that a strict binary division is actually not tenable because across languages, some verbs fail to behave consistently with respect to certain diagnostics. For example, it has been shown that, with standard diagnostics, certain verbs such as *last, stink, bleed, die,* etc can be classified as unaccusative in one language, unergative in a different language (Rosen, 1984; Zaenen, 1988, among many others). This situation is referred to as unaccusativity mismatches. New proposals that specifically focus on these problems have also been made (e.g., Sorace, 2000, below).

2.1 The Connection of Syntax and Semantics in SI

Following the initial theoretical discussions about the connection between syntactic diagnostics and their semantic underpinnings, various semantic factors were suggested. These involve directed change and internal/external causation (Levin & Rappaport-Hovav, 1995), inferable eventual position or state (Lieber & Baayen, 1997), telicity and controllability (Zaenen, 1993), and locomotion (see Randall, 2007; Alexiadou et al., 2004, and, Aranovich, 2007, and McFadden, 2007 for reviews). Some researchers have suggested that syntax has no role in SI. For example, van Valin (1990), focusing on Italian, Georgian, and Achenese, proposed that SI is best characterized in terms of Aksionsart and volitionality. Kishimoto (1996) suggested that volitionality is the semantic parameter that largely determines unaccusative/unergative distinction in Japanese.

Auxiliary selection is among the most reliable syntactic diagnostics proposed for SI. This refers to the auxiliary selection properties of languages that have two perfect auxiliaries corresponding to *be* and *have* in English. In Romance and Germanic languages such as Italian, Dutch, German, and to a lesser extent French, the equivalents of *be* (*essere*, *zijn*, *sein*, *etre*) tend to be selected by unaccusative predicates while the equivalents of *have* (*avere*, *haben*, *hebben*, *avoir*) tend to be selected by unergative predicates (Burzio, 1986; Zaenen, 1993; Keller, 2000; Legendre, 2007, among others). In (1a–b) the situation is illustrated in French (F), German (G) and Italian (I). (Examples are from Legendre, 2007).

(1) a. *Maria a travaillé (F)/hat gearbeitet (G)/ha lavorato (I).*
'Maria worked.'
b. *Maria est venue (F)/ist gekommen (G)/é venuta (I).*
'Maria came.'

Van Valin (1990) and Zaenen (1993) discuss auxiliary selection as a manifestation of the semantic property of telicity. Hence in Dutch, *zijn*-taking verbs are by and large telic, *hebben*-taking verbs are atelic.

Impersonal passivization is another diagnostic that seems applicable to a wide range of languages and used by a number of authors, e.g. Perlmutter (1978), Hoekstra and Mulder (1990), Keller (2000). This construction is predicted to be grammatical with unergative clauses but not with unaccusative clauses. Zaenen (1993) notes that impersonal passivization is controlled by the semantic notion of protagonist control in Dutch; therefore incompatibility of examples such as *bleed* with impersonal passivization is attributed to the fact that *bleed* is not a protagonist control verb. Levin and Rappaport-Hovav (1995:141) take impersonal passivization as an unaccusativity diagnostic but take its sensitivity to protagonist control as a necessary but an insufficient condition for unergative verbs to allow it. In other words, only unergative verbs will be found in this construction, though not all of them.

Refinements of the UH have also been proposed. Most notably, Sorace (2000) argued that the variation attested across languages (as well as within the dialects of a single language) is orderly, and that there are a number of cut-off points to which verb classes can be sensitive.

Sorace's work on (monadic) intransitive verbs is built on variation in the perfective auxiliary selection of verbs in Romance and Germanic languages and called Auxiliary Selection Hierarchy (ASH). She demonstrates that the variation is based on a hierarchy of thematic and aspectual specification of the verbs (viz., telicity and agentivity) and that it is a function of the position of a verb on the hierarchy. Verbs with a high degree of aspectual and thematic specification occupy the extreme ends; variable verbs occupy the middle position, reflecting the decreasing degree of aspectual specification. Both cross-dialectally and across languages, these verbs may be used either with unaccusative or unergative syntax.[3] The ASH therefore is a descriptive statement considering auxiliary selection as a property characterized by both syntax and semantics, as originally viewed by the UH.

We now turn to Turkish, which lacks perfective auxiliaries. A number of other syntactic diagnostics, reviewed below, have been proposed but unlike auxiliaries in other languages, these are not obligatory constructions in Turkish. In addition, the semantic properties underlying the proposed diagnostics have not been studied extensively. Therefore, Turkish presents a particular challenge for any study about SI.

[3] The claim that the two notions of ASH lie within a single dimensional hierarchy has been questioned by Randall (2007). The ASH has also been criticized since it does not explain the reason why a certain language shows the pattern it does (McFadden, 2007).

3 Diagnostics for SI in Turkish

Just as other languages, intransitive verbs in Turkish are sensitive to a set of syntactic environments, summarized below.

3.1 The –*ArAk* Construction

One of the diagnostics is the –*ArAk* construction, which is an adverbial clause formed with the root verb plus the morpheme –*ArAk* (Özkaragöz, 1986). In a Turkish clause which involves the verbal suffix –*ArAk*, both the controller (the complement verb) and the target (the matrix verb) have to be either unaccusative or unergative. In addition, both the controlled and the target have to be the final (surface) subjects of the clause. The examples below contain sentences where both the controller and the target verbs are unaccusative (2) or both are unergative (3). The examples also contain ungrammatical sentences where the controller verb is unergative whereas the target verb is unaccusative (4), and those in which the controller verb is unaccusative whereas the target verb is unergative (5). (Examples are from Özkaragöz, 1986).

(2) *Hasan [kol-u kana -y -arak] acı çek -ti.*
arm-POSS bleed-GL-*ArAk* suffer -PST
'Hasan, while his arm bled, suffered.'

(3) *Kız [(top) oyna-y -arak] şarkı söyle-di.*
girl ball play-GL-*ArAk* sing -PST
'The girl, while playing (ball), sang.'

(4) * *Kız [(top) oyna-y -arak] kay-dı.*
girl ball play-GL-*ArAk* slip -PST
'The girl, while playing (ball), slipped.'

(5) * *Kız [kayak kay-arak] düş-tü.*
girl ski-*ArAk* fall-PST
'The girl, while skiing, fell.'

3.2 Double Causatives

Double construction is allowed with unaccusative verbs but not with unergatives, as shown in (6) and (7) below (Özkaragöz, 1986).

(6) *Sema Turhan-a çiçeğ-i sol- dur -t -tu.*
-DAT flower-ACC fade-CAUS-CAUS-PST

‘Sema made Turhan cause the flower to fade.’

(7) ** Ben Turhan-a Sema-yı koş-tur -t -t -um*
I -DAT -ACC run-CAUS-CAUS-PST-1sg
‘I made Turhan make Sema run.’

3.3 Gerund Constructions

The gerund constructions *–Irken* ‘while’ and *–IncE* ‘when’ are further diagnostics. The former denotes simultaneous action and the latter denotes consecutive action. Unergative verbs are predicted to be compatible with the *–Irken* construction, whereas unaccusatives are predicted to be compatible with the *–IncE* construction, as shown in (8) and (9).[4]

(8) *Adam çalış-ırken esne-di.*
man work-*Irken* yawn-PAST.3per.sg
‘The man yawned while working.’

(9) *Atlet takıl-ınca düş-tü.*
athlete trip-*IncE* fall-PAST.3per.sg
‘The athlete when tripped fell.’

3.4 The Suffix *–Ik*

It has also been suggested that the derivational suffix *–Ik*, used for deriving adjectives from verbs, is compatible with unaccusatives but not with unergatives, as shown in (10) and (11).

(10) *bat-ık gemi*
sink-Ik ship
the sunk ship

(11) **çalış-ık adam*
work-Ik man
the worked man

3.5 The *–mIş* Participle

The past participle marker *–mIş*, which is used for deriving adjectives from verbs has been proposed as yet another diagnostic. The suffix *–mIş* forms participles with transitive and intransitive verbs, as well as passivized verbs. The basic requirement for the acceptability of the *–mIş* participle is the existence of an internal argument in the clause. In well-formed *–mIş* participles, the modified noun must be the external argument of a transitive verb (e.g., *anne* ‘mother’ in [12]), or the internal argument of a passivized verb (e.g., *borç* ‘debt’ in [13]). The internal argument of a transitive verb is not allowed as the modified noun as illustrated in (14).

(12) *Çocuğu-n-u bırak-mış anne*
Child-POSS-ACC leave-*mIş* mother
‘a/the mother who left her children’

(13) *Öde-n-miş borç*
pay-PASS-*mIş* debt
‘the paid debt’

(14) **Öde-miş borç*
pay-*mIş* debt
*‘the pay debt’

As expected, the adjectives formed by intransitive verbs and the *–mIş* participle is more acceptable with unaccusatives compared to unergatives, as shown in (15) and (16).

(15) *sol-muş/ karar-mış çiçek*
wilt/ blacken -*mIş* flower
‘The wilted/blackened flower’

(16) **sıçra-mış/ yüz-müş/ bağır-mış çocuk*
jump/ swim/ shout -*mIş* child
‘The jumped/ swum/ shouted child’

3.6 Impersonal Passivization

Impersonal passivization, used as a diagnostic to single out unergatives by some researchers, appears usable for Turkish as well. In Turkish, impersonal passives carry the phonologically conditioned passive suffix marker, *-Il*, accompanied by an indefinite human interpretation and a resistance to agentive by-phrases. It has been suggested that the tense in which the verb appears affects the acceptability of impersonal passives: when the verb is in the aorist, the implicit subject has an arbitrary interpretation, i.e. either a generic or existential interpretation. On the other hand, in those cases when the verb is in past tense, the implicit subject has a referential meaning, namely a first person plural reading. It was therefore suggested that impersonal passivization is a proper diagnostic environment only in the past tense, which was also adopted in the present study (Nakipoğlu-Demiralp, 2001, cf. Sezer, 1991). (17) and (18) exemplify

[4] Examples in sections 3.3 and 3.4 are from Nakipoğlu (1998).

impersonal passivization with the verb in the past tense.

(17) *Burada koşuldu.*
Here run-PASS-PST
'There was running here.' (existential interpretation)

(18) *??Bu yetimhanede büyündü.*
This orphanage-LOC grow-PASS-PST
'It was grown in this orphanage.'

The diagnostics summarized above do not always pick out the same verbs in Turkish. For example, most diagnostics will fare well with the verbs *düş-* 'fall', *gel-* 'come', *gir-* 'enter' (with a human subject) just as well as impersonal passivization. In other words, these verbs are unaccusative according to most diagnostics and unergative according to impersonal passivization. The opposite of this situation also holds. The stative verb *devam et-* 'continue' is bad or marginally acceptable with most diagnostics as well as impersonal passivization.

The conclusion is that in Turkish, acceptability judgments with the proposed diagnostic environments do not yield a clear distinction between unaccusative and unergative verbs. In addition, it is not clear which semantic properties these diagnostics are correlated with. The model described below is expected to provide some answers to these issues. It is based on native speaker judgments but it goes beyond them by computationally showing that there are correspondences between semantic notions and syntactic manifestations of SI in Turkish. The model is presented below.

4 The Model

This study employs feed-forward artificial neural networks with a backpropagation algorithm as computational models for the analysis of unaccusative/unergative distinction in Turkish.

4.1 Artificial Neural Networks and Learning Paradigms

An artificial neural network (ANN) is a computational model that can be used as a non-linear statistical data modeling tool. ANNs are generally used for deriving a function from observations, in applications where the data are complex and it is difficult to devise a relationship between observations and outputs by hand. ANNs are characterized by interconnected group of artificial neurons, namely nodes. An ANN generally has three major layers of nodes: a single input layer, a single or multiple hidden layers, and a single output layer. In a feedforward ANN, the outputs from all the nodes go to succeeding but not preceding layers.

There are three major learning paradigms that are used for training ANNs: supervised learning, unsupervised learning, and reinforcement learning. A backpropagation algorithm is a supervised learning method which is used for teaching a neural network how to perform a specific task. Accordingly, a feed-forward ANN with a backpropagation algorithm is a computational tool that models the relationship between observations and output by employing supervised learning method (see Hertz et al., 1991; Anderson & Rosenfeld, 1988, among many others for ANNs). The following section presents how such an ANN is used for analyzing unaccusative/unergative distinction in Turkish.

4.2 The Analysis

Two feed-forward ANNs with a backpropagation algorithm were developed for the analysis. Both models had a single input layer, a single hidden layer, and a single output layer of nodes. Both models had a single output node, which represents the binary status of a given verb as unaccusative (0) or unergative (1). The number of nodes in the hidden layer was variable (see below for a discussion of network parameters).

The difference between the two models was the design of the input layer. The first model (henceforth, the diagnostics model DIAG) took diagnostics as input nodes, whereas the second model (henceforth, the semantic parameters model SEMANP) took semantic parameters as input nodes, as presented in detail below.

The Diagnostics Model (DIAG): Binary acceptability values of the phrases or sentences formed by the syntactic diagnostics constituted the input nodes for the network (see above for the SI diagnostics). Each syntactic diagnostic provided a binary value (either 0 or 1) to one of the input nodes. For example, consider the *–mIş* participle as one of the syntactic diagnostics for SI in Turkish. As discussed above, the *–mIş* participle forms acceptable adjectival phrases with unaccusative verbs (e.g., *sol-* 'wilt') but not with the unergative verbs (e.g., *sıçra-* 'jump, leap'), as shown in (19) and (20) below.

(19) *sol-muş çiçek-ler*
wilt-*mIş* flower-PLU
'wilted flowers'

(20) **sıçra-mış sporcu-lar*
jump-*mIş* sportsman-PLU
'jumped sportsmen'

Accordingly, for the verb *sol-* 'wilt', the *–mIş* participle diagnostic provides the value *1* with one of the input nodes, whereas for the verb *sıçra-* 'jump, leap' it provides the value *0* with the corresponding input node. In this way, the syntactic diagnostics constituted an input pattern with eight members for each verb.[5] The construction of an input pattern is exemplified in (21) for the unergative verb *konuş-* 'talk'.

(21) *A sample input pattern for DIAG.*

a. **Adam konuşarak kızardı.* :0
The man talk-*ArAk* blush-PST
'The man blushed by talking.'

b. *Adam konuşarak yürüdü.* :1
The man talk-*Arak* walk-PST
'The man walked by talking.'

c. **Adam kadına çocuğu konuşturttu.* :0
'The man made the woman have the boy talked.'

d. *Adam konuşurken yürüdü.* :1
The man talk-*Irken* walk-PST
'The man walked while talking.'

e. *Adam konuşunca yürüdü.* :1
The man speak-*IncE* walk-PST
'The man walked when he talked.'

f. **Konuş-uk adam* :0
Talk-*Ik* man
'The talked man'

g. **Konuş-muş adam* :0
Talk-*mIş* man
'The talked man'

h. *Törende konuşuldu.* :1
Ceremony-LOC talk-PASS
'It was spoken in the ceremony.'

[5] One of the syntactic diagnostics (the gerund suffix *–ArAk*) involves two verbs (i.e., the target and the matrix verb). Therefore, two sentences/phrases were formed–one with unaccusatives and the other with unergatives–which provided two binary values with the input pattern.

Accordingly, the input pattern for the verb *konuş-* 'talk' is schematically shown below.

0	1	0	1	1	0	0	1

The Semantic Parameters Model (SEMANP): The input nodes for the SEMANP network were constituted by four binary values that represented the status of four semantic parameters (telicity, volitionality, dynamicity, and directed motion) for each verb. Each semantic parameter provided a binary value (either *0* or *1*) to one of the input nodes. The value of the input nodes were determined by applying the following tests for the relevant semantic aspects: (1) *in/for an hour* test for telicity (e.g. the phrase *to talk *in/for an hour* shows that the verb *talk* is atelic whereas the phrase *to wilt in/*for an hour* shows that the verb *wilt* is telic), (2) *on purpose* test for volitionality, (3) *hala-* 'still' test for dynamicity, (4) and the dative test (i.e., acceptability of adding a dative term to the verb) for directed motion.

The construction of an input pattern for SEMANP is exemplified in (22) for the unaccusative verb *sol-* 'wilt'.

(22) *A sample input pattern for DIAG.*

a. Telic :1
b. Non-volitional :0
c. Non-dynamic :0
d. No directed motion :0

Accordingly, the input pattern for the verb *sol-* 'wilt' is schematically shown below.

1	0	0	0

4.3 The Training Phase

The network was trained by providing patterns for 52 verbs that are recognized as unaccusatives in the SI literature or placed closer to the unaccusative end rather than the unergative end of the Auxiliary Selection Hierarchy (ASH, Sorace, 2000); and 52 verbs that are recognized to be unaccusative in the SI literature or placed closer to the unergative end rather than the unaccusative end of the ASH. As a result, a total of 104 input patterns, each composed of eight nodes, were used to train the DIAG model and 104 input patterns, each composed of four nodes, were used to train the SEMANP model. The single output node was set to *0* if the verb with the given input pattern was unaccusative and it was set to *1* if the verb was unergative. Supervised learning method was used, as employed by the backpropagation algorithm.

One hidden layer with a variable number of hidden units was used (see below for the analysis of model parameters). Sigmoid activation function, shown in (23), was used for modeling the activation function.

(23) $$f(x) = \frac{1}{1+e^{-x}}$$

The number of maximum iterations per epoch was set to 20. The system sensitivity was defined by a global variable (ε=0.01) which decided whether the loops in the code converge or not.

4.4 The Test Phase

The DIAG and SEMANP models were tested by providing the following input patterns:

Group A: five verbs that are either recognized as unaccusatives in the SI literature or placed closer to the unaccusative end rather than the unergative end of the ASH.

Group B: Five verbs that are either recognized as unergatives in the SI literature or placed closer to the unergative end rather than the unaccusative end of the ASH.

Group C: Three verbs that are reported to exhibit variable behavior within the ASH.

After the training, the networks provided the binary outputs for the test verbs, which showed whether a test verb was unaccusative or unergative according to the models.

5 Results

The results are presented in the two sections below, separately for the DIAG model and for the SEMANP model.

5.1 The DIAG Model

After the training of the network and the optimization of the number of hidden units and the learning rate, the DIAG model classified all verbs in Group A as unaccusatives. The model also classified all Group-B verbs as unergatives. Finally, the model categorized three Group-C verbs that were reported to show variable behavior (*kana-* ‘bleed’, *parla-* ‘shine’ and *üşü-* ‘be, feel cold’) as unaccusative verbs in Turkish.

The distribution of weights after the training showed that the *–mIş* participle received the highest weight, which indicates that the *–mIş* participle is the most reliable diagnostics for analyzing unaccusative/unergative distinction in Turkish.

5.2 The SEMANP Model

The SEMANP model classified two of the Group-A verbs (namely, *gir-* ‘enter’ and *yetiş-* ‘grow’) as unaccusatives and the three remaining verbs (*dur-* ‘remain, stay’, *kal-* ‘stall, stay, and *varol-* ‘exist’) as unergatives. The model also classified four of five Group-B verbs (*gül-* ‘laugh’, *sırıt-* ‘grin’, *söylen-* ‘mutter’, *yakın-* ‘complain’) as unergatives and the remaining verb (*yüz-* ‘swim’) as unaccusative. Finally, the model categorized three Group-C verbs (*kana-* ‘bleed’, *parla-* ‘shine’ and *üşü-* ‘be, feel cold’) as unaccusative verbs in Turkish.

The distribution of weights after the training showed that among the four semantic parameters that were selected in this study, telicity received the highest weight, which indicates that unaccusative and unergative verbs are most sensitive to the telicity aspect of the verb in Turkish.

5.3 Evaluation of Model Parameters

Four model design parameters, their initial values and acceptable ranges after optimization are discussed below.

The number of hidden units: The number of hidden layers was set to 1 as a non-variable design parameter of the network. The initial number of hidden units was set to 3. Keeping the learning rate (η=-0.25) and the momentum term (λ=0.25) constant, the number of hidden units was adjusted and the behavior of the network was observed. The analyses showed that the optimum range for the number of hidden units was between 2 and 6.

The learning rate: The learning rate was initially set to η=-0.25. Keeping the number of hidden units (hidden_size=3) and the momentum term (λ=0.25) constant, adjusting the learning rate between η=-0.005 and η=-0.9 did not have an effect on the results.

The momentum term: The momentum term was set to λ=0.25 initially. Keeping the number of hidden units (hidden_size=3) and the learning rate (η=-0.25) constant, adjusting the momentum term between λ=0.01 and λ=1.0 did not have an effect on the results. However, the system did not converge to a solution for the momentum term equal to and greater than λ=1.0.

6 Discussion

A major finding of the suggested model is that the predictions of the two models are compatible with the UH (Perlmutter, 1978) in that they divide most intransitive verbs into two, as expected. Furthermore, the differences between the decisions of the diagnostics-based DIAG model and the semantic-parameters-based SEMANP model reflect a reported finding in the unaccusativity literature, i.e., the tests used to differentiate between unaccusatives and unergatives do not uniformly delegate all verbs to the same classes (the solution of why such mismatches occur in Turkish is beyond the scope of this study, see Sorace, 2000; Randall, 2007, for some suggestions). More specifically, the three Group-A verbs that were predicted as unaccusative by the DIAG model and unergative by the SEMANP model (*dur*- 'remain, stay', *kal*- 'stall, stay, and *varol*- 'exist') are stative verbs, which are known to show inconsistent behavior in the literature and classified as variable-behavior verbs by Sorace (2000). An unexpected finding is the Group-B verb (*yüz*- 'swim'), which is predicted as unergative by the DIAG model and unaccusative by the SEMANP model. This seems to reflect the role of semantic parameters other than telicity (namely, dynamicity and directed motion) in Turkish. The remaining nine verbs of thirteen tested verbs were predicted to be of the same type (either unaccusative or unergative) by both models.

Another finding of the model is the alignment between the most weighted syntactic diagnostics for unaccusative/unergative distinction in Turkish, namely the *–mIş* participle which received the highest weight after the training, and the most weighted semantic parameter, namely telicity.

7 Conclusion and Future Research

This study contributes to our understanding of the distinction in several respects.

Firstly, it proposes a novel computational approach that tackles the unaccusative/unergative distinction in Turkish. The model confirms that a split between unaccusative and unergative verbs indeed exists in Turkish but that the division is not clear-cut. The model suggests that certain verbs (e.g., stative verbs) behave inconsistently, as mentioned in most accounts in the literature. Moreover, the model reflects a correspondence between syntactic diagnostics and semantics, which supports the view that unaccusativity is semantically determined and syntactically manifested (Permutter, 1978, Levin & Rappaport-Hovav, 1995). Since this approach uses relevant language-dependent features, it is particularly applicable to languages that lack explicit syntactic diagnostics of SI.

The computational approach is based on the connectionist paradigm which employs feed-forward artificial neural networks with a back-propagation algorithm. There are several dimensions in which the model will further be developed. First, the reliability of input node values will be strengthened by conducting acceptability judgment experiments with native speakers, and the training of the model will be improved by increasing the number of verbs used for training. Acceptability judgments are influenced not only by verbs but also by other constituents in clauses or sentences; therefore the input data will be improved to involve different senses of verbs under various sentential constructions. Second, alternative classifiers, such as decision trees and naïve Bayes, as well as the classifiers that use discretized weights may provide more informative accounts of the findings of SI in Turkish. These alternatives will be investigated in further studies.

Acknowledgements We thank Cem Bozşahin and three anonymous reviewers for their helpful comments and suggestions. All remaining errors are ours.

References

Alexiadou, Artemis, Anagnostopoulou, Elena, and Everaert, Martin (eds.). 2004. *The unaccusativity puzzle: Explorations of the syntax-lexicon interface.* Oxford University Press.

Anderson, James, and Rosenfeld, Edward. 1988. *Neurocomputing: Foundations of research.* MIT Press.

Aranovich, Raúl (ed.). 2007. *Split auxiliary systems.* Amsterdam/Philadelphia: John Benjamins.

Burzio, Luigi. 1986. *Italian syntax: A Government-Binding approach*. Dordrecht: Reidel.

Hertz, John, Krogh, Anders, and Palmer, Richard G. 1991. *Introduction to the theory of neural computation*. Addison-Wesley, Massachusetts.

Hirakawa, Makiko. 1999. L2 acquisition of Japanese unaccusative verbs by speakers of English and Chinese. In *The acquisition of japanese as a second language*, ed. Kazue Kanno, 89-113. Amsterdam: John Benjamins.

Hoekstra, Teun, and Mulder, René. 1990. Unergatives as copular verbs: Locational and existential predication. *The Linguistic Review* 7: 1-79.

Keller, Frank. 2000. Gradience in grammar: Experimental and computational aspects of degrees of grammaticality. University of Edinburgh dissertation.

Kishimoto, Hideki. 1996. Split intransitivity in Japanese and the Unaccusative Hypothesis. *Language* 72: 248-286.

Legendre, Géraldine. 2007. On the typology of auxiliary selection. *Lingua* 117: 1522-1540.

Levin, Beth, and Rappaport-Hovav, Malka. 1995. *Unaccusativity at the syntax-lexical semantics interface*. MIT Press.

Lieber, R., and Baayen, H. 1997. A semantic principle of auxiliary selection in Dutch. *Natural Language and Linguistic Theory* 15: 789-845.

McFadden, Thomas. 2007. Auxiliary selection. *Language and Linguistics Compass* 1/6: 674-708.

Mithun, Marianne. 1991. Active/agentive case marking and its motivations. *Language* 67: 510-546.

Nakipoğlu, Mine. 1998. Split intransitivity and the syntax-semantics interface in Turkish, Minneapolis: University of Minnesota dissertation.

Nakipoğlu-Demiralp, Mine. 2001. The referential properties of the implicit arguments of impersonal passive constructions. In *The verb in Turkish*, ed. E. Eser Taylan, 129-150, John Benjamins.

Oshita, Hiroyuki. 1997. The unaccusative trap: L2 acquisition of English intransitive verbs. University of Southern California dissertation.

Özkaragöz, İnci Z. 1986. The relational structure of Turkish syntax, San Diego: University of California dissertation.

Perlmutter, David M. 1978. Impersonal passives and the Unaccusative Hypothesis. In *Proceedings of the Fourth Annual Meeting of the Berkeley Linguistics Society*, ed. Farrell Ackerman et al., 157-189.

Randall, J. 2007. Parameterized auxiliary selection: A fine-grained interaction of features and linking rules. In *Split auxiliary systems*, ed. Raúl Aranovich, 207-236. John Benjamins.

Richa, Srishti. 2008. *Unaccusativity, unergativity and the causative alternation in Hindi: A minimalist analysis*. New Delhi: Jawaharlal Nehru University dissertation.

Rosen, Carol G. 1984. The interface between semantic roles and initial grammatical relations. In *Studies in relational grammar*, eds. Carol G. Rosen and David M. Perlmutter. University of Chicago Press.

Sezer, Engin. 1991. *Issues in Turkish syntax*. Cambridge: Harvard University dissertation.

Sorace, Antonella. 2000. Gradients in auxiliary selection with intransitive verbs. *Language* 76: 859-890.

Sorace, Antonella, and Shomura, Yoko. 2001. Lexical constraints on the acquisition of split intransitivity. Evidence from L2 Japanese. *Studies in Second Language Acquisition* 23: 247-278.

van Valin, Robert D., Jr. 1990. Semantic parameters of split transitivity. *Language* 66: 221-260.

Zaenen, Annie. 1988. Unaccusative verbs in Dutch and the syntax-semantics interface. CSLI Reports 88-123, CSLI, Stanford University.

Zaenen, Annie. 1993. Unaccusativity in Dutch: Integrating syntax and lexical semantics. In *Semantics and the lexicon*, ed. James Pustejovsky. Dordrecht: Kluwer Academic Publishers.