An Unaccusative Categorization of ‘Swarm’ Type Verbs in Spanish: An Application of SAS® to Linguistic Research

Roberto Mayoral Hernández, University of Alabama at Birmingham
Lindsey Chen, National Taiwan Normal University

ABSTRACT
The universal classification of intransitive verbs into two classes - unaccusatives and unergatives - plays a central role in modern linguistics as it provides information about the word order and other fundamental properties of verbs. In particular, there exists a controversial type of verbs, the ‘swarm’ type, which has been classified as both unaccusative (Perlmutter 1978) and unergative (Levin & Rappaport 1995). In this paper, SAS® statistical procedures were employed to provide evidence from Spanish that shows that the ‘swarm’ type should be categorized as unaccusative rather than unergative. In order to determine the best classification, data from the corpus CREA (Corpus de Referencia del Español Actual) was extracted and via SAS® Proc Freq, results from chi-square tests show that these verbs display a high percentage of postverbal subjects while unergatives have a higher percentage of preverbal subjects.

INTRODUCTION
There exist two basic types of verbs: transitive and intransitive. Transitive verbs are those that take two arguments - a subject (S) and a direct object (DO); see (1a). Typically, the subject of transitive verbs has agentive properties. It is an agent because it functions as the initiator or causer of the action described by the verb. The direct object, on the other hand, has patient properties. It is the theme or patient that is affected by the main action described by the verb (see Dowty 1991 for further discussion on the properties of the agent and patient thematic-roles). By contrast, intransitive verbs only select a single argument, which is the subject. The intransitive type can be subdivided further into two subgroups: unaccusatives and unergatives (Perlmutter 1978, Hale & Kaiser 2002, Levin 1993). Grosso modo, unergatives and unaccusatives can be differentiated because unaccusatives select a subject with theme/affected properties (1b), while unergatives take agentive subjects (1c).

(1) a. John (S) eats apples (DO).
   b. The train (S) has already arrived.
   c. John (S) works hard.

The distinction between unaccusatives and unergatives has led some (cf. Perlmutter 1978) to state that the subject of unaccusative verbs is an underlying object. In spite of the abundant research on the subject (cf. Levin & Rappaport 1995 and the references therein), there is no universally accepted test for unaccusativity or unergativity. This is particularly true for languages that do not mark the distinction with specific morphology such as Spanish or English. Some types of verbs, those that imply a change of state or location, have generally been classified as prototypical unaccusatives rather uncontroversially due to the existence of overt grammatical markings that identify them crosslinguistically. However, there are some controversial types of verbs that have received diverse classifications in the literature. One of these verbs is the so called ‘swarm’ type. Verbs such as swarm have been both described as unaccusative (Perlmutter 1978) and unergative (Levin & Rappaport 1995). This paper provides evidence from Spanish, via statistical analyses with SAS, that supports an unaccusative classification of the ‘swarm’ type when they appear in certain constructions, by applying the subject position test developed in Mayoral Hernández (2004). In particular, we show that the ‘swarm’ type verbs are clearly associated with a higher percentage of postverbal subjects, which has traditionally been related to unaccusative structures (Torrego 1989). These results support a structural (or constructional) view of unaccusativity.

INTRANSITIVITY: UNACCUSATIVE VS. UNERGATIVE
In this section we provide a general overview of the features that differentiate unergative and unaccusative verbs, using crosslinguistic data. Perlmutter 1978 divided intransitive verbs into two subgroups: unaccusative and unergative. Both classes have distinct semantic and syntactic features which make them different from one another. Without going into much detail, Perlmutter associated unergative verbs with a notion of volition or causation, and unaccusative verbs with affected arguments. The differences existing between these two classes of verbs have also explicit syntactic consequences. Specifically, this author provides examples to show that while unergative verbs can appear in the impersonal passive construction in Dutch (2), unaccusatives are not allowed (3).
Impersonal passive with unergative verb in Dutch.
Er wordt hier veel geskied.
“It is skied here a lot”

Active sentence (a) and impersonal passive (b) with unaccusative verb in Dutch
Dat blok hout heeft goed gebrand.
“That block of wood burned well”

*Er werdt door dat blok hout goed gebrand.
“By that block of wood it is burned well”

Later, Burzio 1981, 1986 described the properties of unaccusative verbs in Italian, and came to the realization that unaccusative and unergative verbs can be differentiated by using different syntactic tests:

Unaccusative verbs select the auxiliary essere “be” for periphrastic forms.
Unergative verbs select the auxiliary avere “have” for periphrastic forms.
Only unaccusative verbs accept ne-cliticization.
The participle appearing in periphrastic forms of unaccusative verbs agrees with the subject, while there is no overt agreement with the subject of unergative verbs.

Therefore, in languages that have two auxiliaries used to form periphrastic tenses, unaccusative verbs will select the auxiliary be, while unergative verbs will select have. The following sentences show examples from Dutch and Italian.

Both languages use the auxiliary be, zijn and essere respectively, in unaccusative constructions, while unergative verbs select the auxiliary have, hebben and avere respectively.

The Italian partitive clitic ne can be used to substitute for the direct object of a transitive verb, as in (7), or the subject of an unaccusative verb, such as in (8). However, it cannot substitute for the subject of an unergative verb (9). This test seems to suggest that the subject of an unaccusative verb behaves in the same way as the object of a transitive verb, since the quantified object of a transitive verb can also be replaced by the pronominal form ne, but not its subject.

In Italian, the past participle of periphrastic forms agrees with the subject in gender and number when the auxiliary be is used, i.e. with unaccusative verbs (10a). Unergative verbs, however, disallow this possibility (10b).
(10) a. Maria è arrivata (Fem, sg)  
    "Maria has arrived"

b. Maria ha telefonato (Mas, sg)  
    "Maria has telephoned"

The literature dealing with unaccusativity is very extensive and it is beyond the scope of this paper to carry out a complete review of it. Nevertheless, it is crucial to provide a description of Spanish intransitive verbs in order to determine the right classification of the ‘swarm’ type verbs. As in Italian, there have been several tests aimed at differentiating unergatives from unaccusative verbs in Spanish. However, some of the most uncontroversial tests, such as auxiliary selection, cannot be applied to Spanish, because only *haber* “have” can be used as an auxiliary in periphrastic forms. Mendikoetxea 1999 and the references therein are a good source for a detailed description of Spanish intransitive verbs.

UNACCUSATIVITY IN SPANISH

The lack of overt unaccusative morphology in Spanish complicates the task of distinguishing verb types in Spanish. Although the literature proposes several grammatical tests (absolute participial clauses, the causative alternation, bare postverbal subjects,…), they all present abundant exceptions. Mayoral Hernández 2006 proposes a statistical test which shows that there is a statistically significant difference between unergative and unaccusative verbs in Spanish as far as subject position is concerned (table 1).

<table>
<thead>
<tr>
<th>Subject Position</th>
<th>Type of Verb</th>
<th>Unergative</th>
<th>Unaccusative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postverbal</td>
<td>Count</td>
<td>32</td>
<td>62</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Column %</td>
<td>21.5%</td>
<td>45.6%</td>
<td>33.0%</td>
</tr>
<tr>
<td>Preverbal</td>
<td>Count</td>
<td>117</td>
<td>74</td>
<td>191</td>
</tr>
<tr>
<td></td>
<td>Column %</td>
<td>78.5%</td>
<td>54.4%</td>
<td>67.0%</td>
</tr>
<tr>
<td>Total</td>
<td>Count</td>
<td>149</td>
<td>136</td>
<td>285</td>
</tr>
<tr>
<td></td>
<td>Column %</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

In particular, while unergative verbs show a clear preference for preverbal subjects, the subjects of unaccusatives present a similar distribution of preverbal and postverbal positions.

In this paper we will adopt this procedure to identify the best classification of the ‘swarm’ type verbs.

STATISTICAL ANALYSIS WITH SAS

This experiment takes as a point of departure the corpus used in Mayoral Hernández 2006 & 2008. It includes a total of 2141 sentences that contain the following verbs: *empujar* “push”, representing transitive verbs; *trabajar* “work”, prototypical unergative; *llegar* “arrive”, typical unaccusative; and the ‘swarm’ type verbs *resonar* “resound”, *abundar* “abound”, *arder* “burn”, *brillar* “shine”, *hervir* “boil” and *pulular* “swarm”. Data obtained from the corpus CREA were coded and then analyzed using SAS. The variables taken into consideration are subject position (preverbal and postverbal) and type of verb (unergative, transitive, unaccusative, ‘swarm’ type). The independent variable is the subject position and the dependent variable is the type of verb. Below is a sample code that creates the dataset for the subject distribution in unergative and swarm verbs:

```sas
Data verbs1;
Input position $ type $ count @@;
Datalines;
Post swarm 835 post unerg 32
Pre swarm 854 pre unerg 118
;
proc freq data=verbs1;
table position*type/expected chisq nocol
norow nopercent;
weight count;
title 'Subject distribution in unergative and swarm verbs';
run;
```
RESULTS
The comparison of unergative and ‘swarm’ type verbs indicates that there is a statistically significant difference between them, as far as subject position is concerned. Table 2 shows that the ‘swarm’ type verbs have an almost even distribution across preverbal and postverbal domains, while unergative prefer preverbal subjects.

Table 2: Subject distribution in unergative and ‘swarm’ verbs (p < .0001)

<table>
<thead>
<tr>
<th>position</th>
<th>type</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>swarm</td>
<td>unerg</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>pre</td>
<td>854</td>
<td>0</td>
<td>854</td>
<td></td>
</tr>
<tr>
<td></td>
<td>794.34</td>
<td>69.657</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>32</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29.39</td>
<td>2.6101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>post</td>
<td>0</td>
<td>118</td>
<td>118</td>
<td></td>
</tr>
<tr>
<td></td>
<td>108.36</td>
<td>9.6248</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>150</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1689</td>
<td>1686</td>
<td>3375</td>
<td></td>
</tr>
</tbody>
</table>

Conversely, the comparison of unaccusative and the ‘swarm’ type verbs shows that they have very similar percentages of preverbal and postverbal subjects, always around 50%. In particular, the ‘swarm’ verbs have a 49.4% of postverbal subjects, compared to a 54.8% of postverbal subjects with unaccusative verbs.

Assuming that the subject position test constitutes an accurate means to identify verb types in Spanish, we can postulate that the ‘swarm’ type verbs should be categorized as unaccusative. This position had already been adopted in the literature by Perlmutter 1978, and seems to contradict Levin & Rappaport’s 1995 classification. If we push the results of this experiment a little further, one might be tempted to defend a lexical categorization of verbs. This would imply that verb types are determined in the lexicon, where they receive certain structural properties. However, we will not adopt such thesis here. Instead, we show that a more detailed observation of the results points toward a different direction.

LOCATIVE ALTERNATION
All ‘swarm’ type verbs are known to enter an intransitive version of the locative alternation. This construction is characterized by the fact that its arguments can either appear as a subject (S) or a prepositional object (PO), as in example (11).

(11) a. The girl’s chanting (S) resounds in the hallways (PO).
    b. The hallways (S) resound with the girl’s chanting (PO).

There is some evidence that suggests that the locative alternation construction has unaccusative properties. This would imply that any verb used in this construction, whether typically unaccusative or not, will acquire unaccusative properties. Because of this, we annotated the type of construction where swarm type verbs appear, in order to determine if they behave similarly whether or not they alternate. Our results suggest that these verbs are pretty stable
across constructions, and they either show a preference for postverbal subjects or no preference at all. However, there is one exception. The verb *brillar* "shine" seems to prefer preverbal subjects whenever it appears in sentences with no locative alternation (table 3), while postverbal subjects become more frequent in alternating constructions. In this sense, even if the statistics indicate a significant difference between the verb *brillar* and a typically unergative verb, as table 3 shows, both of them clearly prefer preverbal subjects (70.3% of preverbal subjects with non-alternating *brillar*, compared to 78.7% of preverbal subjects with unergative verbs). The results are different when the verb *brillar* appears in the locative alternation construction, where the percentage of preverbal subjects becomes only 43.3%, which is expected only from an unaccusative verb.

Table 3: Subject position by verb type: non-alternating *brillar* and unergative verbs (p < .0001)

<table>
<thead>
<tr>
<th>Position</th>
<th>Expected</th>
<th>Actual</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>brillar</td>
<td>unerg</td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>0</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>22.062</td>
<td>9.9379</td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>118</td>
<td>118</td>
<td>234</td>
</tr>
<tr>
<td></td>
<td>36.646</td>
<td>30.745</td>
<td></td>
</tr>
<tr>
<td>post</td>
<td>99</td>
<td>0</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>88.256</td>
<td>30.745</td>
<td></td>
</tr>
<tr>
<td>pre</td>
<td>234</td>
<td>0</td>
<td>234</td>
</tr>
<tr>
<td></td>
<td>161.88</td>
<td>72.671</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>383</td>
<td>150</td>
<td>483</td>
</tr>
</tbody>
</table>

The fact that a given verb may show so different subject distributions suggests that an *a priori* division of verbs into distinct types might not be descriptively adequate in some contexts. In the case of the "swarm" type, the results indicate that they are not necessarily unaccusative by themselves. However, when they appear in the locative alternation construction they always show a higher percentage of postverbal subjects without exception. What transpires from this data is that the locative alternation construction has always unaccusative properties. The verbs inserted in it do not need to be necessarily unaccusative to enter this construction. This is not a novel claim. For example, it is generally accepted that directed motion constructions in languages such as Dutch have unaccusative properties, even though some typically unergative verbs (manner of motion verbs) can also appear in them (See Zubizarreta & Oh 1997 for a more detailed discussion on this topic).

CONCLUSION
This paper has shown that it is possible to identify verb types by using statistical techniques and the software SAS. The experiments presented here indicate that a controversial type of verbs, the "swarm" type, show clear unaccusative properties when they occur in the locative alternation construction. In order to reach this conclusion, we have applied the subject position test, which is based on the assumption that there is a measurable relationship between subject position and verb types in Spanish. The results indicate that "swarm" type pattern together with unaccusative verbs as far as subject position is concerned, since they show no preference for either preverbal or postverbal subjects. However, the control of the type of construction where these verbs appear discourages an *a priori* lexical categorization of these verbs as unaccusative, since they show a variable behavior that is directly related

The results indicate that "swarm" type pattern together with unaccusative verbs as far as subject position is concerned, since they show no preference for either preverbal or postverbal subjects. However, the control of the type of construction where these verbs appear discourages an *a priori* lexical categorization of these verbs as unaccusative, since they show a variable behavior that is directly related...
to the construction where they appear. Instead, our results indicate that whenever these verbs are inserted in the locative alternation construction, they will show unaccusative properties without exceptions, which indicates that it is the locative alternation construction the one that always presents unaccusative properties. This paper thus supports a constructional view of unaccusativity.

REFERENCES

REAL ACADEMIA ESPAÑOLA: (CREA) data base [online]. *Corpus de referencia del español actual*. http://www.rae.es

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Roberto Mayoral Hernández
University of Alabama-Birmingham
Department of Foreign Languages and Literatures

Lindsey Chen
National Taiwan Normal University

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
Other brand and product names are trademarks of their respective companies.