To be or not *to be*: A corpus-based study of unaccusative verbs and auxiliary selection

Master's Thesis

Presented to

The Faculty of the Graduate School of Arts and Sciences

Brandeis University

Department of Computer Science

James Pustejovsky, Advisor

In Partial Fulfillment

of the Requirements for

Master's Degree

By

Richard A. Brutti Jr.

May 2012

ABSTRACT

To be or not *to be*: A corpus-based study of unaccusative verbs and auxiliary selection

A thesis presented to the Department of Computer Science

Graduate School of Arts and Sciences
Brandeis University
Waltham, Massachusetts
Richard A. Brutti Jr.

Since the introduction of the *Unaccusative Hypothesis* (Perlmutter, 1978), there have been many further attempts to explain the mechanisms behind the division in intransitive verbs. This paper aims to analyze and test some of theories of unaccusativity using computational linguistic tools. Specifically, I focus on verbs that exhibit *split intransitivity*, that is, verbs that can appear in both unaccusative and unergative constructions, and in determining the distinguishing features that make this alternation possible. Many formal linguistic theories of unaccusativity involve the interplay of semantic roles and temporal event markers, both of which can be analyzed using statistical computational linguistic tools, including semantic role labelers, semantic parses, and automatic event classification. I use auxiliary verb selection as a surface-level indicator of unaccusativity in Italian and Dutch, and

test various classes of verbs extracted from the Europarl corpus (Koehn, 2005). Additionally, I provide some historical background for the evolution of this distinction, and analyze how my results fit into the larger theoretical framework.

Contents

List of Tables

CHAPTER 1

The Unaccusative/Unergative Distinction

"To be, or not to be, that is the question"

William Shakespeare, *Hamlet*, Act 3 Scene 1

1.1. What does it mean to be *unaccusative*?

As first articulated by Perlmutter (Perlmutter, 1978), the *Unaccusative Hypothesis* (UH) describes a division of intransitive verbs. The main idea is that the surface-level subject of certain intransitive verbs, the so-called *unaccusatives*, is really an object in its underlying representation (whatever that may be). The surface-level subject of *unergatives* is underlyingly a subject. The UH encompasses fundamental questions on the relationship between syntax and semantics: are verbs unaccusative because of the syntactic structure in which they appear? or do they appear in certain syntactic constructions due to semantic constraints?

This paper will attempt to answer some of these questions by using computational techniques. Specifically, I will build a sentence-level classifier for unaccusativity, with the feature selection based on theoretical and historical descriptions. In the following chapters, I will provide an overview, while gleaning useful features for my analysis along the way. Using NLP tools such as syntactic parsers and semantic role labelers, I aim to quantify the phenomenon of unaccusativity.

The subject of unaccusative verbs is underlyingly at some level an object (the terms subject and object are not used in the original UH, but I will do so here)

(Perlmutter, 1978). The subject of unergative verbs is initially a subject, and remains so on surface level. As presented by Burzio in the Government-and-Binding framework, there is a binary distinction of intransitive verbs (Burzio, 1986):

(1) John telephones.

(2) John arrives.

Since the UH, unaccusativity has been described in many of the world's languages, but there is not a single cross-linguistic diagnostic criterion. Fully syntactic analyses omit the possibility that semantics influence unaccusativity, while semantic approaches do not permit a syntactic motivation for unaccusativity (Levin & Rappaport Hovav, 1995). However, fairly reliable diagnostics have been borne out from both approaches.

In this thesis, I consider the syntactic realizations (i.e., auxiliary verb choice) to be an effect of the semantics, and am specifically interested in single verbs that can appear in both unaccusative and unergative constructions by analyzing the lexical and compositional features that allow this alternation to occur. This will be discussed further in Chapter 4.

Unaccusativity is not a unified phenomenon; even individual languages show internal gradience (Sorace, 2004). Specifically, I am focusing my study on verbs that Italian and Dutch verbs that pair with both auxiliaries when intransitive. To borrow a pun (intended or not) from Perlmutter, I will analyze the 'perfect' auxiliary (as well as the pluperfect auxiliary, future perfect auxiliary et al.).

1.2. Grounding Unaccusativity

Burzio describes unaccusativity in terms of constituent movement that takes place from underlying form to the surface level. Burzio identifies a class of AVB/BV alternations, where V represents a verb, and A and B are noun phrases, in which single verbs can encode different relationships between constituents:

(3) L'artiglieria affondò due navi nemiche.
the artillery sank two ships enemy
'The artillery sank two enemy ships.'

(4) Due navi nemiche affondarono.
two ships enemy sank
'Two enemy ships sank.'

The above Italian examples show a transitive/unaccusative distinction, a *paired* unaccusative. This contrasts with *unpaired* unaccusatives like *exist*, *arrive*, etc. that have no corresponding transitive (Perlmutter, 1989).

The following Dutch example shows a paired unaccusative with an unergative alternant (van Hout, 2004):

(5) John is in vijf minuten naar huis gelopen.
John is in five minutes to home walked
'John walked home in five minutes.'

(6) John heeft urenlang gelopen.
John has for hours walked
'John walked for hours.'

These examples illustrate auxiliary verb choice as a reliable surface level diagnostic for unaccusativity. Dutch unaccusatives select *zijn* 'be' and unergatives select *hebben* 'have'. Italian parallels this, with unaccusatives selecting *essere* 'be' and unergatives *avere* 'have'.

As is already becoming clear, there is much variation in unaccusative verbs cross-linguistically. Verbs that are unaccusative in one language may not be so in another. As illustrated by the Dutch examples above, the "underlying object" hypothesis is insufficent for explaining the unaccusative/unergative dichotomy in all cases.

The same verb can have unaccusative and unergative alternants in Italian as well (examples from (Pustejovsky & Busa, 1995)):

(7) Giovanni è corso a casa.
Giovanni BE ran to house
'Giovanni ran home.'

(8) Giovanni ha corso nel bosco.
Giovanni HAVE ran in the woods
'Giovanni ran in the woods.'

1.3. Unaccusativity Diagnostics in Italian

1.3.1. Ne-cliticization

Various phenomenon in Italian are coindicated with unaccusativity. One of the first recognized unaccusativity diagnostics was the distribution of the Italian clitic *ne*. *Ne* 'of it/them' typically appears as a direct object in transitive and unaccusative constructions, but not with unergatives (Burzio, 1986):

(9) a. Ne arriveranno molti.
of them will arrive many
'Many of them will arrive.'

b. * Ne telefoneranno molti.
of them will call many

Since *ne* appears with unaccusatives, it should be coindicated with auxiliary choice; *ne* can appear with verbs that take *essere* and not with verbs that take *avere*. While this is largely the case, Lonzi (1985) provides examples of *avere* verbs permitting *ne*-cliticization, but only in the unmarked simple present tense (Levin & Rappaport Hovav, 1995):

(10) a. * Ne ha camminato tanta, di gente, su quei marciapiedi.
of them HAVE walked many of people on those sidewalks

b. Ne cammina tanta, di gente, su quei marciapiedi.
of them walk many of people on those sidewalks
'Many people walk on those sidewalks.'

Evidence of this type is enough to launch a corpus study of *ne*-cliticization. Based on semantic approaches to unaccusativity, this may be due to the fact that the present tense sentence describes an event that has not been completed yet. More discussion of this will be forthcoming.

It certainly brings the status of *ne* as an unaccusative diagnostic under some doubt. At worst, it is still a very strong indicator of unaccusativity; and I will consider it a feature in my corpus study.

1.4. Unaccusativity Diagnostics in Dutch

1.4.1. Passivization

One of the first unaccusative diagnostics recognized in Dutch was the ability to form of impersonal passives (Perlmutter, 1978). Unergatives may be passivized in a *there*-construction, while unergatives may not (Zaenen, 1988):

(11) a. De jongens werkten.
the boys worked
'The boys worked/were working.'

b. Er werd (door de jongens) gewerkt.
there was (by the boys) worked

'There was worked by the boys.'

(12) a. De jongens vielen.
the boys fell
'The boys fell.'

b. * Er werd door de jongens gevallen.
there was by the boys fell

Since passive constructions lose the subject role, unaccusatives cannot be passivized, while unergatives can (Alexiadou, Anagnostopoulou, & Everaert, 2004). Impersonal passives are a strong indicator of unaccusativity, but not a foolproof test. Zaenen cites some exceptions from Perlmutter (1978) in which some verbs that take *hebben* (an indicator of unergativity) cannot be passivized (Zaenen, 1988):

(13) a. Dat hout heeft goed gebrand.
that wood HAVE well burned
'That wood has burned well.'

b. * Er werd door dat hout goed gebrand.
there was by that wood well burned

As with *ne*, this could be the subject of an interesting corpus study. This would require a more detailed analysis of Dutch lemmas and the various environments in which they can appear.

1.4.2. Prenominal Past Participles

Dutch unaccusatives permit a past participle to appear prenominally and have an active meaning, while both unaccusatives and unergatives allow the same usage with a present (gerundive) participle (Zaenen, 1988):

(14) a. * De gewerkte man.
the worked man

‘The worked man.’

b. De werkende man.
the working man
‘The working man.’

(15) a. Het gevallen blad.
the fell leaf
‘The fallen leaf.’

b. Het vallende blad.
the falling leaf
‘The falling leaf.’

Again, this is not a perfect fit, but there are only 3 verbs that take zijn and cannot appear in this construction - *blijven* ‘remain’, *groeien* ‘grow’, and *gaan* ‘go’ (Zaenen, 1988).

1.5. Cross-linguistic Diagnostics

1.5.1. Resultatives

Resultative phrases have been considered a reliable unaccusativity diagnostic across many languages (Levin & Rappaport Hovav, 1995). Levin and Rappaport Hovav illustrate with examples from English. These phrases can appear as objects of a transitive verb (Levin & Rappaport Hovav, 1995):

(16) Woolite safely soaks all your fine washables clean. (ad)

This construction cannot appear in verbs that do not have an object, so unergatives cannot appear with resultatives:

(17) * Dora shouted hoarse.

Unergative verbs can appear in resultatives that describe the state reached by the referent of the NP, resulting in the action that was carried out by the verb (Levin & Rappaport Hovav, 1995):

(18) The hockey coach skated the team ragged.

A third type has "the NP following the unergative verb is a nonsubcategorized inalienably possessed NP (generally denoting a body part) where the possessor is coreferential with the subject of the verb" (Levin & Rappaport Hovav, 1995):

(19) a. Dance your cares away. (*Fraggle Rock*)

b. * Dance your cares.

Unaccusatives can appear with resultatives that are based on their grammatical subjects (Levin & Rappaport Hovav, 1995):

(20) The gate swung shut.

Since the surface subject of unaccusatives is considered an underlying object, the consistency is maintained. Unaccusatives cannot appear with resultatives as in the constructions outlined above (Levin & Rappaport Hovav, 1995):

(21) a. * The snow melted itself slushy.

b. * The snow melted the road slushy.

These resultative phrases indicate *accomplishments* from verbs that usually indicate *activities* (Levin & Rappaport Hovav, 1995); accomplishments are analyzed as having an internal *event structure* with a separate activity and resulting *state* (Pustejovsky, 1995). Certain verbs, like *build*, are inherently accomplishments with two internal events (a process and a state). Resultatives describe a change of state in verbs that are not necessarily change-of-state verbs when appearing alone

(Levin & Rappaport Hovav, 1995). Event structure will play an important role in our further discussion of unaccusativity.

1.5.2. Locative Inversion

Levin & Rappaport Hovav present the case of locative inversion as an unaccusativity diagnostic (1995). It is usually found with typical unaccusatives (a), like verbs of appearance and existence. Unergatives usually cannot appear in this construction (b) :

(22) Over her shoulder appeared the head of Jenny's mother.

(23) * In the nursery smiled half a dozen newborn babies.

They ultimately determine that the preverbal PP is a subject, at some level, and that this phenomenon is largely a function of discourse and not unaccusativity.

Tortora posits an implicit locative in a class of Italian unaccusatives to explain certain constructions with marked word order. Verbs like *arrivare* form a GOAL-entailing subclass, can only appear in verb-subject unmarked context if there is a telic interpretation (Tortora, 2001):

(24) a. L'aereo è sceso (sulla pista) in 5 minuti.
the airplane BE descended on the runway in 5 minutes
'The airplane descended (onto the runway) in 5 minutes.

b. È sceso lo Spitfire (*per 5 minuti).
BE descended the Spitfire (for 5 minutes)
'The Spitfire descended (*for 5 minutes).'

Using the null locative as a feature in the computational study would prove be very difficult, but the pairing of prepositional phrases above will prove to be consistent with semantic notions of event type, discussed in Chapter 3.

1.6. Why is this interesting?

Why is this interesting? Unaccusativity has been documented as a "psychologically real" phenomenon (Sorace, 1993) and there is even "neurological support" for the unaccusative/unergative distinction (Shetreet, Friedmann, & Hadar, 2010). Studies have been performed in both second-language learners of languages with unaccusativity as well as native speakers of languages with unaccusativity, and there are indeed differences in how each category of verb is processed mentally.

Work out of Tel Aviv University has provided neurological support for the unaccusative hypothesis. While I will leave the details to the experts, the authors found that the cortical representation of unaccusatives differs from the representation of unergatives, and even located approximate brain areas where each is processed (Shetreet et al., 2010).

While Shetreet et al. largely avoided verbs with unaccusative and unergative alternants, they are at the center of (or more accurately, at the bottom of) Sorace's *Unaccusativity Hierarchy.* She conducted studies of French and Italian non-native speakers of Italian and French respectively. French maintains the same Romance auxiliary split, but the distribution is not the same. The results show that unaccusativity is an internally consistent system that it is a linguistically and psychologically real phenomenon (Sorace, 1993).

1.7. Omissions from this study - Si et al.

Si is one of the most complex aspects of the Italian language, and has received its own treatment many times over for all of its roles (D'Alessandro, 2001). There are many phenomena related to *si* and unaccusativity, but for the sake of a computational study of unaccusatives, *si* cannot provide much information as feature because all verbs in *si* constructions take *essere*.

While we are on the subject, there are a other constructions that we will necessarily ignore; passives have been discussed as an unaccusativity diagnostic, but will not be used as a feature for the classifier. Additionally, I will not look at auxiliary verb selection in regards to raising verbs, and verbs like *must*, *can*, *want*, etc. that take clausal complements.

CHAPTER 2

The Haves and the Have-nots

2.1. Descriptive Approaches to Auxiliary Selection

As stated, I am considering auxiliary verb selection to be the surface-level reflex of unaccusativity. As such, this chapter will examine historical and descriptive approaches to auxiliary verb selection in Italian and Dutch, and thereby unaccusativity in not so few words. The notion of subjects and objects behaving differently goes back to at least Sapir in 1917 (Perlmutter, 1989).

2.1.1. Italian

Italian verb conjugations require an auxiliary verb in various tenses and aspects, including the *passato prossimo* (recent past), *trapassato prossimo*, (pluperfect), *futuro anteriore* (future perfect), *trapassato* (historical past), *condizionale passato* (conditional preterite), *congiuntivo passato* (subunctive preterite), and the *congiuntivo trapassato* (subjunctive pluperfect). The auxiliary expresses the person but is combined with a participle which carries gender and number information when *essere* is the auxiliary (Napolitano & Devine, 1979):

(25) Io sono salita sull'albero.
I BE climbed up the tree
'I climbed the tree.'

(26) Tu hai salito le scale.
you HAVE climbed the stairs
'You climbed the stairs.'

In traditional grammars, the distinction between auxiliaries is usually described as a split along the lines of transitivity; intransitive verbs take *essere* and transitive verbs take *avere*. However, grammar books are quick to point out exceptions, especially movement. A more advanced Italian grammar is slightly more specific, "*Usiamo essere con: i verbi di movimento [e] i verbi che indicano un divenire.*" (We use *essere* with: verbs of movement [and] verbs that indicate becoming) (Napolitano & Devine, 1979). These parameters are more formal, however there is still a fair list of exceptions.

Descriptive grammars of Italian seem to be hinting at the unaccusative/ unergative distinction. Certain Italian intransitives take *avere*; the list of so-called exceptions is expanded to "*verbi intransitivi che indicano un'attivitá del corpo o della mente*" (intransitive verbs that indicate an activity of the body or of the mind), as well as "*verbi intransitivi indicanti movimento non direzionale*" (intransitive verbs indicating non-directional movement) (Moretti & Orvieto, 1979).

2.1.2. Dutch

The situation is similar in Dutch. The auxiliaries *zijn* and *hebben* are used in various compound tenses, including the *voltooid tegenwoordige tijd* (perfect), *voltooid verleden tijd* (pluperfect), *voltooid tegenwoordig toekomende tijd* (future perfect), and the *voltooid verleden toekomende tijd* (conditional perfect). The auxiliary carries person information and is combined with the participle of the main verb. The verb participle does not change with number and gender, (Donaldson, 1997):

(27) a. Ik ben gevallen.
I BE fallen
'I have fallen.'

b. Ik heb mijn paraplu vergeten.
I HAVE my umbrella forgot
'I have forgotten my umbrella.'

2.1.3. Dutch Exceptions

Descriptions of the *zijn/hebben* distinction are more convoluted, and there is generally less philological material on Old Dutch than Old Italian. Broadly, the vast majority of transitive verbs take *hebben*.

Certain verbs (including *dansen* 'dance' and *fietsen* 'cycle') take *zijn* if they describe a " motion to or from a particular place". The same verbs take *hebben* when used in a more descriptive, but also intransitive, manner (Donaldson, 1997). Few transitive verbs take *zijn* (including *beginnen* 'begin', *naderen* 'approach', and *aankomen* 'lose weight').

2.2. Historical Approaches to Auxiliary Selection

2.2.1. Romance

Historically, *avere* and *essere* come from the Classical Latin *habere* and *esse*. *Habere* and *esse* are transitive and intransitive respectively, so their Italian counterparts are relatively consistent (Maiden, 1995). However, *essere* is the auxiliary for all reflexive verbs. Normal rules of agreement do not apply for Italian reflexive unaccusative verbs, while unergatives must agree (D'Alessandro, 2001):

(28) Si è arrivati.
si BE arrived
'People/we arrived at home.'

(29) Si è telefonato.
si BE called

‘People/we called.’

We will ignore these issues in our study, since the auxiliary verb use is invariable.

As a transitive verb, *habere* requires two arguments, and the subject may have the semantic role of agent or experiencer (Maiden, 1995):

(30) Giovanni accende la radio.
Giovanni turns on the radio.
‘Giovanni turns on the radio.’

(31) Giovanni sente il freddo.
Giovanni feels the cold
‘Giovanni feels the cold.’

Historically, there is also a strong connection between the subject of Italian transitive verbs and the locative role, since the preposition *da* (from) is used in passive constructions to indicate the agent (Maiden, 1995):

(32) Il libro è letto da Giovanni.
The book BE read from/by Giovanni
‘The book is read by Giovanni.’

In modern Italian, it can be argued that certain intransitive *avere* verbs have an underlying locative as the subject (Maiden, 1995). This analysis fits nicely with Levin & Rappaport’s discussion of internal vs. external verbs of motion and locative inversion clauses (Levin & Rappaport Hovav, 1995).

The grammaticalization of *avere* to intransitives likely occurred in late Latin, but it is impossible to determine if the change was based on syntactic or semantic analogy.

Latin *esse* is intransitive, and its Italian reflex, *essere*, selects one argument for its subject. The historical case of *esse* and intransitives is straightforward (Maiden, 1995). As discussed above, it is the auxiliary for all modern reflexives. A large corpus of 14th century Italian, Dante's *Commedia*, there are instances of *avere* as an auxiliary in reflexive constructions (Calchini, 2011), reflecting a gradual process of grammaticalization.

(33) Fatto v'avete dio d'oro e d'argento (*Inferno XIX*)
Made to-you HAVE god of gold and of silver
'You've made yourselves a god of gold and silver' (Mandelbaum)

2.2.2. Germanic

Philological treatments of the auxiliary split in Germanic languages describe it as modeled on the Latin distinction (Benveniste, 1966). The Old High German work *Tatian* (c.825 A.D.) exhibits the *sein/haben* distinction, likely as a result of imitating Latin work (Priebsch, 1948).

Philological accounts of Germanic describe *haben* as the auxiliary for verbs that take an object, with only a few exceptions (*folgen*, 'follow'). Priebsch's thorough account discusses verbs that do not indicate a change of state taking *haben. Haben* is used with 'cessative' verbs (atelic), and interestingly for this account, Priebsch notes that in Modern German, "it is customary in the north (of Germany) to make a clear distinction between *ich habe geschwommen* 'I have been swimming' (non-terminate occurrence) and *ich bin ùber den Fluss geschwommen* (terminate occurrence)." The distinction is largely the same in Dutch, but Priebsch notes that "'it is curious that Dutch says even with a direct object *ik ben't vergeten.*'

In what may be the earliest precursor to Generative Lexicon-style event structure, Priebsch further elucidates the distinction in (§37) (Priebsch, 1948):

> The combination of the preterite of *haben* or *sein* with the past participle originally indicated a state of things already existent in the past...Subsequently the combination indicated the process rather than the resultant state, viz. a process which has already taken place before an occurrence in the past, so that the pluperfect is an ante-preterite (*he had just gone when I came*).

2.3. Semantic Roles and Event Types

I have been alluding to notions of semantic roles and event types without any formal definitions of these terms. Vendler's classification of events (1967) is as follows (Van Valin Jr, 1990):

(34) a. STATES: *know*, *be broken*, *have*, *believe*, *like*

b. ACHIEVEMENTS: *learn*, *break* (intr.), *die*, *arrive*, *notice*

c. ACCOMPLISHMENTS: *teach*, *break* (tr.), *kill*, *eat a piece of pizza*

d. ACTIVITIES: *run*, *dance*, *swim*, *eat pizza*, *squeak*

2.3.1. Event Types

Vendler's event types have been further classified for the sake of automatic event classification algorithms (Zarcone & Lenci, 2008):

> Accomplishment and achievement predicates are *telic*, while states and activities are *atelic*. On the other hand, states, activities and accomplishments are *durative*, and achievements are *non durative*.

I will further explore event classification in Chapters 3 and 4.

2.3.2. Semantic roles

This is not the place to explore the nuances of thematic relation definitions (or to discuss their merits), as it has been done so elsewhere. For computational purposes, they have been largely subsumed by PropBank (Palmer, Gildea, & Kingsbury, 2005). For the unaccusative/unergative distinction, notions of agent and patient/theme are sufficient to contrast the deliberate performer of an action as opposed to the undergoer of an action. PropBank will use ARG0 and ARG1 here.

Historical descriptions predating these theories do not use these notions in quite the same way, but the terms appeal to intuitive notions, so it is not necessary here to 'translate' Priebsch or Napolitano's descriptions into modern terminology.

2.4. UA/UE in Related Languages

Even in closely related languages like Dutch and German, the borderline between split intransitive verbs varies (Randall, 2004):

(35) a. Dutch: John heeft urenlang op de tafel gedanst.

b. German: John hat stundenlang auf dem Tisch getanzt.
John HAVE for hours on the table danced
'John danced on the table for hours.'

(36) a. Dutch: John heeft urenlang door de zaal rondgedanst.

b. German: John ist stundenlang durch den Saal herumgetanzt.
John HAVE(NL)/BE(DE) for hours around the room danced
'John danced around the room for hours.'

Unaccusativity is inconsistent within Romance languages as well. The following French sentences are from (Ruwet, 1972) as discussed in (Alexiadou et al., 2004), and the Italian sentences appear in (Pustejovsky & Busa, 1995):

(37) a. French: L'ennemi a coulé le bateau.

b. Italian: I nemici hanno affondato la nave.

the enemies HAVE sank the boat

'The enemies sank the boat.'

(38) a. French: Le bateau a coulé.

b. Italian: La nave è affondata.

the boat HAVE(FR)/BE(IT) sank

'The boat sank.'

It is not my intention to examine unaccusativity universally, however these minimal pairs illustrate the subtleties of the distinction, even in closely related languages. Evidence of this kind is enough to show that there will be no cross-linguistic diagnostic for unaccusativity, and that the most informative features will vary by language.

To further complicate the matter, Sorace posits that an evolution is taking place, at least in Romance, towards HAVE as the auxiliary in all cases. The movement is largely complete in Spanish, in which, like English, all perfect tenses take HAVE. French *avoir* is used in many categories where Italian *essere* is the auxiliary. However, French change of location verbs (*aller* 'go', *venir* 'come') take *être*, again suggesting that verbs of directed motion are central to the unaccusativity phenomenon (Sorace, 1993).

CHAPTER 3

Theoretical Approaches to the UA/UE Distinction

In this chapter, I will trace the major linguistic theories of unaccusativity and auxiliary selection, and attempt to weave together their underlying similarities. These proposals have been discussed in depth by linguists more qualified than myself, so I will provide a brief overview and discussion on how they fit into my computational analysis.

3.1. Syntactic Accounts

3.1.1. Perlmutter

In the Relational Grammar framework, verbs are transitive iff they have a *2-arc* (analogous to a direct object). The UH distinguishes two types of *strata* for intransitive verbs, unaccusative and unergative. "A stratum is *unergative* if and only if it contains a 1-arc and no 2-arc," and "*unaccusative* if and only if it contains a 2-arc and no 1-arc" (Perlmutter, 1989):

(39) a. *soffrire*:[1 (2)] 'suffer'

b. *affogare*:[(1) 2] 'drown'

Without delving too deeply into the strata of relational grammar, Perlmutter basically claims that verbs like *correre* that can appear with either auxiliary have two distinct entries, initially unergative if appearing with *avere* and initially unaccusative if appearing with *essere*.

3.1.2. Burzio

Burzio adapts and expands on Perlmutter's work in the Principles and Parameters system (Burzio, 1986). *Ne*-cliticization is the defining test for unaccusativity, which is represented as a $[_{NP}$ e] V NP structure. The empty subject has no thematic role, and the underlying direct object does not get assigned accusative case. The null subject is filled by the object, and gets nominative case. All unaccusatives take *essere*, and all other verbs take *avere*. It is a wide-ranging and purely syntactic account, but does not really consider verbs like *correre* that have unergative and unaccusative alternations (Centineo, 1996). This is the fundamental work on the syntax of unaccusatives, but Burzio now believes that semantics is where the more substantive issues remain (personal communication).

3.2. Semantic Accounts

3.2.1. Manner of Motion Verbs

I have been alluding to semantic factors of unaccusativity since discussing resultative constructions in Chapter 1. Levin & Rappaport Hovav's unaccusative *Verbs of inherently directed motion* (such as *come*, *go*, *arrive*) cannot appear with resultatives. Also called *Run* verbs, verbs of this category are unergative when appearing by themselves, and unaccusative when used with directional phrases. When these verbs appear with resultatives, the resultative acts as a time delimeter; they are telic and unaccusative, as seen in 5 and 6.

Levin & Rappaport Hovav offer the most comprehensive account of unaccusativity, and they analyze a wide range of syntactic and semantic proposals before offering their own semantically motivated, but syntactically realized theory.

It is not my intention to summarize their account, but to use features that I deem salient for a computational approach.

3.2.2. Chierchia

For Chierchia's semantic approach, unaccusatives are special type of reflexives, specifically of causatives (Chierchia, 2004). Causatives are represented as (a) in Montague's notation. (b) is the representation in Chierchia's richer semantic system which includes entities, properties, and propositions as separate types (I will omit a longer description here):

(40) a. λ x λ y $\exists$ P[CAUSE(P(y), α(x)]

b. λ x $^{\cap}$ λ y $\exists$ β[CAUSE($^{\cup}\beta$(y), $^{\cup}\alpha$(x)]

He proposes a type-shifting operation for reflexivization, *R*, in the spirit of Partee, such that the meaning of the unaccusative *affondare* 'sink' becomes some property or state of the boat that causes it(self) to sink. *R* is of type $\langle\langle e, \pi\rangle \rightarrow \pi\rangle$, where e is the type of entities and π is the type of properties. Again, (a) is the Montague approximation, and (b) is the representation in the expanded property theory:

(41) a. λ x [wash(x)(x)]

b. $^{\cup}$[R(wash)](x) $\leftrightarrow^{\cup}$[wash(x)](x)

In the example below, the above causative is abbreviated as C(α), and the unaccusative *affondare* is represented as a property of the boat that causes it to sink:

(42) a. affondare$_{IV}$ = R(affondare$_{TV}$) (=R(C(α)x)]

La barca è affondata. ‘The boat sank.’

b. $^{\cup}$ R(affondare$_{TV}$)](the boat)

Of course, there are unaccusatives that do not have a transitive alternant. They are still represented as R(C(α)) which provides an account for the ‘instability’ associated with unaccusatives - we have seen the same verbs vary cross-linguistically, dialectically, and historically. The C(α) formation is necessarily internal to the verb. The verb *crescere* ‘grow’ is only unaccusative in Italian, but in some Italian dialects (Chierchia, 2004) and in Dante (Calchini, 2011), there are transitive uses, explained by the presence of C(α):

(43) a. I pomodori sono cresciuti.
the tomatoes BE grown
‘The tomatoes grew.’

b. * Gianni ha cresciuto pomodori.
Gianni HAVE grown tomatoes

c. e che più volte v’ha cresciuta doglia? *Inferno IX*
and that more times to-you HAVE grown pain
‘and which so often added to your hurts?’ (Mandelbaum)

I would not expect to find such uses in the Europarl corpus, but C(α) opens the door for metaphoric or novel uses of unpaired unaccusatives.

Since Chierchia posits unaccusatives to be reflexives of causative verbs, this also relates to the telic/atelic hypotheses which are central to unaccusativity. Due to their internal event structure, causatives necessarily involve bringing about a state (Pustejovsky, 1995). If a causative verb is not stative, it will have a telic interpretation. Additionally, non-stative unaccusative verbs will be achievements or accomplishments. As we will see, telicity has been posited as the determining factor for Dutch unaccusativity (van Hout, 2004).

Dowty identified the now canonical test for telicity; telic verbs take for-PPs, and atelic verbs take in-PPs (Dowty, 1979):

(44) a. John pushed the card for an hour.

b. ?? John drew a circle for an hour.

c. ?? John pushed a cart in a minute.

d. John drew a circle in a minute.

Our favorite unaccusative/unergative alternant verb, *run*, follows the same pattern (Chierchia, 2004):

(45) a. ?? Gianni è corso in giardino per un'ora.
Gianni BE ran in garden for an hour

b. Gianni ha corso in giardino per un'ora.
Gianni HAVE ran in garden for an hour

c. Gianni è corso in giardino in un minuto.
Gianni BE ran in garden in a minute

d. Gianni ha corso in giardino in un minuto.
Gianni HAVE ran in garden in a minute

Run is usually interpreted atelically, as an activity or a process. In the above examples, when *correre* is unergative it combines with the telic for-PPs. This parallels the Dutch examples in Chapter 1, and will prove to be a salient feature for the classifier.

3.2.3. van Hout

Van Hout's work in Dutch identifies telicity as the lone factor for determining unaccusativity. This example was used in Chapter 1:

(5) John heeft urenlang gelopen.
John HAVE for hours walked
'John walked for hours.'

(6) John is in vijf minuten naar huis gelopen.
John BE in five minutes to home walked
'John walked home in five minutes.'

Like *correre* in the discussion of Chierchia's proposal, *lopen* 'walk' is generally atelic. When it combines with a to-PP in (b), the interpretation is telic, it takes *hebben*, and is unergative. The prenominal past participle usage, a strong diagnostic of Dutch unaccusativity (Zaenen, 1988), is consistent with this finding; the atelic interpretation cannot be used as a prenominal past participle, but the telic one can (van Hout, 2004):

(46) a. * de gelopen jongen
the walked boy

b. de naar huis gelopen jongen
the to home walked boy
'the boy who walked home'

3.3. Event Structure

In the Generative Lexicon approach to unaccusativity, as in Chierchia's work, unaccusatives are considered underlying causatives. The unaccusative/unergative distinction rises from the “underspecified” nature of the lexical entry, analogous to Chierchia’s notion of the internal C(α).

A critical component for the study of unaccusatives is the *logical polysemy* of words, uniting what are traditionally considered separate word senses into various instances of one essential meaning of a lexical item (Pustejovsky, 1995).

In this framework, the *Default Causative Paradigm* for causative verbs is represented as:

(47) $[\mathrm{R}(\mathrm{e}_1, \mathrm{x}, \mathrm{y}) \wedge\neg\ \mathrm{P}(\mathrm{e}_1, \mathrm{y}) \wedge \mathrm{P}(\mathrm{e}_2, \mathrm{y}) \wedge\neg\ \mathrm{e}_2 \prec \mathrm{e}_1\,] > \mathit{cause}(\mathrm{e}_1, \mathrm{e}_2)$

Which can be summarized as “If event e_1 happens, then and only then, event e_2 is always produced by it”, as is necessary for causal predicates like *break* or *sink* (Pustejovsky & Busa, 1995). The underspecified nature of verbs like *rompere* ‘break’, which allows for the unaccusative and transitive syntactic realizations, along with the inherent event structure of verbs, can be explained by the notion of *event headedness*. The event structure of transition verbs always contains two subevents, so there are four possible arrangements of event heads (marked with an asterisk):

(48) a. $[^T_e e_1{*} <_\propto e_2\,]$ - build

b. $[^T_e e_1 <_\propto e_2{*}\,]$ - arrive

c. $[^T_e e_1{*} <_\propto e_2{*}\,]$ - give

d. $[^T_e e_1 <_\propto e_2\,]$ - break

We have described unaccusative alternants as “underspecified”, so this analysis will focus on verbs with this structure, like (d). A verb like (a) is an accomplishment, where the focal point is on the action bringing about the state of something being built. Verbs like (b) are achievements, where the focus is on the resulting state, and are unpaired unaccusatives. For the sake of brevity, I will leave further explanation of the other structures to (Pustejovsky & Busa, 1995).

The nature of headedness allows the individual subevents to be modified like grammatical objects. For a verb like *affondare* 'sink', e_1 can be represented as [sink_act(x,y)] (i.e. "x sinks y") and e_2 as [sunk(y)] (i.e. "y is sunk"), with e_1 necessarily coming before e_2 (Pustejovsky, 1995). The unaccusative reading of (38b),"La nave è affondata", grounds the end state of the boat, and the transitive (37b), "I nemici hanno affondato la nave", grounds the agentive predicate. Either way, the boat is grounded:

(49) a. Giovanni ha affondato la nave per incassare l'assicurazione.
Giovanni HAVE sank the ship for collect the insurance
'Giovanni sank the boat to collect insurance.'

b. * La nave è affondata per incassare l'assicurazione.
* the boat BE sank for collect the insurance

As in Chierchia, the modifiers are paramount. The to-PP serves as a function that turns processes into transitions, thereby grounding the right headed event of the resulting state of the run action, and resulting in the unaccusative reading:

(50) Giovanni è corso a casa.
Giovanni BE run to home
'Giovanni ran home.'

Again, as we have seen in Chierchia, not every unaccusative verb is grammatical when used unergatively or transitively. The Italian verb, *camminare* 'walk', only takes *avere* regardless of PP adjuncts (Pustejovsky & Busa, 1995):

(51) a. Gianni ha camminato in giardino per un'ora.

Gianni HAVE walked in garden for an hour

b. * Gianni è camminato a casa.

Gianni BE walked to home

The Dutch verb *lopen* (walk) demonstrates split intransitivity, and as seen in some examples from Dante and Italian dialects, verbs can change their patterns over time and across dialects, so even though Levin and Rappaport Hovav's various categories are not set in stone cross-linguistically, they will serve as useful starting points.

3.4. Conclusions

While I have omitted a number of noteworthy accounts (Van Valin Jr, 1990), (Centineo, 1996), and (Sorace, 2004), I do not believe that they provide additional features for a computational analysis beyond what was discussed in this chapter. An idea that is central to all the semantic accounts is *underspecifity*. The variable behavior of paired unaccusative/unergative verbs are subject to compositional factors external to the lexical item, such as temporal and directional adjuncts. Telicity in usage is not a result of the lexical item, but what it combines with. Unpaired unaccusatives and unergatives do not exhibit the same variability, and auxiliary verb choice is not influenced by adjuncts.

CHAPTER 4

A Computational Approach to the UA/UE Distinction

4.1. Corpus Selection

For my computational investigation into unaccusativity, I used the Europarl Corpus (Koehn, 2005). Europarl consists of nearly 50 million words in the three major European languages that I have been focusing on. While the text is from Parliament proceedings, the corpus is large enough to display a good deal of linguistic variation, and certainly is sufficient for my purposes. Using various verb categories that display a high degree of variability in the unaccusative/unergative alternation, I extracted several thousand sentences in each language pair. Considering auxiliary choice as the most reliable cross-linguistic display of unaccusativity, and labeling my samples as such, I then built a binary sentence-level classifier with a comprehensive featureset to determine the most distinctive features of unaccusative sentences.

4.1.1. Corpus

The Europarl corpus is freely available online, and is packaged by language pair. The corpus is sentence aligned, but there are slight differences between the language pairs, so an Italian-Dutch aligned version, for example, is not available. The Natural Language Toolkit (NLTK) (Loper & Bird, 2002) ships with a small sample subset of the Europarl corpus, accessible through its CorpusReader. I replaced these files with the complete Europarl corpus, enabling me to easily use the entire corpus as an NLTK CorpusReader object. For the sake of keeping the alignments straight, I used four corpora - Italian, Dutch, and the Englishes that were aligned with Italian and Dutch respectively.

From here, I extracted the subset unaccusative/unergative corpus for each language, and built language-specific classifiers with slight variation in sentence extraction and feature set-selection as appropriate.

4.2. Verb Categories

I used six categories of verbs that show high unaccusative/unergative variability cross-linguistically. They are based on categories defined by Levin and Rappaport Hovav, Perlmutter, and Sorace, as well as other variable-behavior verbs that I have identified through research or via native-speaker informants. The following tables contain the Italian and Dutch words respectively. The lists were originally in English, so I have done my best to include as many synonyms as possible (although it was difficult to render words like *skulk* & *skedaddle* in Italian and Dutch).

Verbal categories are imperfect indicators, but since many of these verbs appear with both auxiliaries, they should be coindicated with many of the other features that we have identified. I have tried to choose representative categories of various manner of motion verbs, emission verbs, etc.

4.2.1. Roll Verbs

As discussed in Levin & Rappaport Hovav, the category of *roll* verbs include verbs that do not have an inherently directed manner of motion. When non-agentive, these verbs are largely unaccusative, and many feature transitive and unergative alternants (*the ball rolled down the hill* vs. *the children rolled the ball*) (Levin & Rappaport Hovav, 1995).

Table 4.1. List of *Roll Verbs* in Italian & Dutch

ROLL VERBS
cadere ‘drop/fall’, *dondolare* ‘spin/swing’, *galleggiare* ‘float’, *muovere* ‘move’, *girare* ‘turn/revolve’, *rimbalzare* ‘bounce’, *roteare* ‘twirl’, *rotolare* ‘roll’, *serpeggiare* ‘coil/wind’, *scivolare* ‘glide’, *sdrucciolare*‘slide’, *svoltare* ‘spin’, *turbinare* ‘whirl’, *volteggiare* ‘circle/spin’
draaien ‘turn/rotate’, *drijven*‘drift’, *druipen* ‘drop/drip’, *glijden* ‘glide/slide’, *opscheiten*‘hurry/coil’, *opspringen* ‘bounce’, *rollen* ‘roll’, *ronddraaien* ‘twirl’, *rouleren* ‘rotate’, *schommelen* ‘roll/swing’, *spinnen* ‘spin’, *springen* ‘jump/bounce’, *verhuizen* ‘move’, *vlotten* ‘float’, *wenden* ‘wind/turn’

4.2.2. Run Verbs

Run verbs make up another category in Levin & Rappaport. They are also called *Agentive verbs of manner of motion.* They contain the verbs which we have used as cannonical examples throughout (*correre*, *lopen*, *nuotare*, *zwemmen*, etc.). These verbs describe a motion, but there is not inherent direction (like *come* or *go*) (Levin & Rappaport Hovav, 1995).

4.2.3. Verbs of Sound Emission

According to Levin & Rappaport Hovav, various verbs can have internal or external causes. The subjects of internally caused verbs have some sort of inherent property that brings about the resulting state, but are not necessarily agentive. They do not usually have causative alternants, but do have transitive ones (Levin & Rappaport Hovav, 1995). When combined with directional phrases, *verbs of sound emission* behave like the *run verbs* above; when they are agentive, they cannot become verbs of directed motion (Levin & Rappaport Hovav, 1995):

Table 4.2. List of *Run Verbs* in Italian & Dutch

RUN VERBS
ambiare ‘amble’, *attraversare* ‘trek/cross’, *balzare* ‘jump/bound’, *barcollare* ‘lurch’, *cabotare* ‘coast’, *camminare* ‘walk’, *cappottare* ‘somersault’, *carambolare* ‘carom’, *correre* ‘run’, *filare* ‘spin’, *fuggire* ‘bolt’, *galoppare* ‘gallop’, *gironzolare* ‘saunter’, *inerpicarsi* ‘clamber’, *marciare* ‘march’, *nuotare* ‘swim’, *percorrere* ‘roam’, *saltare* ‘jump’, *saltellare* ‘hop’, *sbrigarsi* ‘hasten’, *scalare* ‘climb’, *schizzare* ‘dart’, *scorrazzare* ‘scamper’, *sdrucciolare* ‘slide’, *sfilare* ‘parade’, *sfrecciare* ‘speed’, *sgambettare* ‘gambol’, *smammare* ‘scram’, *strascicare* ‘shuffle’, *strisciare* ‘crawl’, *svolazzare* ‘flit’, *trascinarsi* ‘traipse’, *trottare* ‘trot’, *vacillare* ‘dodder’, *vagabondare* ‘stray’, *vagare* ‘meander’, *viaggiare* ‘travel/journey’, *volare* ‘fly’, *zigzagare* ‘zigzag’, *zoppicare* ‘hobble’, *zumare* ‘zoom’
afdwalen ‘stray’, *beven* ‘shake/dodder’, *caramboleren* ‘carom’, *chargeren* ‘charge’, *draven* ‘trot’, *duikelen* ‘somersault’, *freewheelen* ‘coast’, *galopperen* ‘gallop’, *hinken* ‘limp’, *hompelen* ‘hobble’, *huppelen* ‘hop’, *klauteren* ‘clamber’, *kletsen* ‘dash’, *klimmen* ‘climb’, *kronkelen* ‘meander’, *kruipen* ‘crawl’, *lopen* ‘run’, *opklimmen* ‘climb’, *opkrassen* ‘scram’, *paraderen* ‘parade’, *reizen* ‘journey/ravel’, *rennen* ‘scamper’, *rondlopen* ‘perambulate’, *rondsluipen* ‘prowl’, *rondspringen* ‘cavort’, *rondzwerven* ‘traipse’, *slaapwandelen* ‘sleepwalk’, *slenteren* ‘saunter’, *springen* ‘bound/jump’, *steigeren* ‘prance’, *stutten* ‘strut’, *suizen* ‘whiz’, *vliegen* ‘fly’, *voorruitschieten* ‘bolt’, *voortslenteren* ‘mosey’, *waggelen* ‘totter’, *wankelen*‘stagger’, *zoemen* ‘buzz/zoom’, *zwemmen* ‘swim’, *zwerven* ‘roam’

(52) * He yelled home. (Cf. He yelled his way home.)

Additionally, here are two examples from Europarl to illustrate the alternation with the same verb, *suonare* ‘sound’:

(53) a. In seguito agli attacchi dell’11 settembre 2001, è
in following to the attacks of the 11 september 2001 BE
improvvisamente suonato il campanello dallarme per tutti
suddenly rang the bell of alarm for all

‘Following the attacks of 11 September 2001, everyone was suddenly given a wake-up call’

b. Quando la campanella ha suonato... ho guardato lo
when the bell HAVE rang HAVE looked the
schermo per vedere cosa stesse succedendo.
screen for see what BE happening

‘When the bell rang...I looked to the screen to see what was happening.’

Table 4.3. List of *Sound Verbs* in Italian & Dutch

VERBS OF SOUND EMISSION
balbettare ‘babble’, *battere* ‘beat’, *chiocciare* ‘squawk’, *chiurlare* ‘hoot’, *cigolare* ‘creak’, *cinguettare* ‘chatter’, *crepitare* ‘crepitate’, *fiorire* ‘bloom’, *fischiare* ‘pipe’, *gemere* ‘moan’, *gorgogliare* ‘gurgle’, *gridare* ‘shriek’, *mormorare* ‘murmur’, *mugghiare* ‘bellow’, *ringhiare* ‘growl’, *risuonare* ‘resound’, *ronzare* ‘buzz’, *sbattere* ‘bang’, *scampanellare* ‘jingle’, *scoppiettare* ‘splutter’, *scricchiolare* ‘crackle’, *sferragliare* ‘clank’, *sibilare* ‘hiss’, *sonicchiare* ‘tootle’, *squillare*‘blare’, *suonare* ‘sound/chime’, *ticchettare* ‘clack’, *tintinnare* ‘chink’, *trillare* ‘trill’, *tuonare* ‘thunder’, *ululare* ‘wail’
babbelen ‘babble’, *blatten* ‘bellow’, *bloeien* ‘bloom’, *brommen* ‘hum’, *brullen* ‘roar’, *donderen* ‘thunder’, *gillen* ‘shriek’, *gorgolen* ‘gurgle’, *grommen* ‘growl’, *jammeren* ‘wail’, *klapperen* ‘chatter’, *kletteren* ‘clang/crackle’, *kraken* ‘creak’, *kreunen* ‘moan’, *loeien* ‘blare’, *luiden* ‘chime’, *mompelen* ‘murmur’, *piepen* ‘squeak’, *rinkelen* ‘jingle’, *schreeuwen* ‘squawk’, *sissen* ‘hiss’, *spinnen* ‘purr’, *toeteren* ‘hoot’, *zoemen* ‘buzz’

4.2.4. Verbs of Bodily Processes

Verbs of bodily processes are internally caused, but do not usually have causative alternants (Levin & Rappaport Hovav, 1995). These verbs vary by language; for example, *blush* is unaccusative in Italian, but unergative in Dutch (Sorace, 2004).

Table 4.4. List of *Body Verbs* in Italian & Dutch

VERBS OF BODILY PROCESSES
arrossire ‘blush’, *dormire* ‘sleep’, *russare* ‘snore’, *ruttare* ‘burp’, *sanguinare* ‘bleed’, *sbadigliare* ‘yawn’, *singhiozzare* ‘hiccup’, *starnutire* ‘sneeze’, *tossire* ‘cough’, *vomitare* ‘vomit’
bloeden ‘bleed’, *bloizen* ‘blush’, *boeren* ‘burp’, *gapen* ‘yawn’, *hikken* ‘hiccup’, *hoesten* ‘cough’, *niezen* ‘sneeze’, *overgeven* ‘vomit’, *slapen* ‘sleep’, *snurken* ‘snore’

4.2.5. Known UA/UE variants

These are verbs that are highly variable cross-linguistically. Some are found in the above categories as well. This set of verbs was not considered a feature in the classifier, since it is so disparate. These verbs are important to include in the building of the UA/UE corpus, and their behavior and cooccurrences with other features will be useful.

Table 4.5. List of Verbs with UA/UE Variation in Italian & Dutch

KNOWN UA/UE VARIANTS
abbassare ‘lower’, *appartenere* ‘belong to’, *affogare* ‘drown’, *affondare* ‘sink’, *atterrare* ‘fell’, *brillare* ‘shine’, *cambiare* ‘change’, *continuare* ‘continue’, *correre* ‘run’, *durare* ‘last’, *fallire* ‘fail’, *fiorire* ‘bloom’, *germogliare* ‘bud’, *guarire* ‘heal’, *inciampare* ‘stumble’, *nuotare* ‘swim’, *mancare* ‘lack’, *marcire* ‘rot’, *rimbombare* ‘reverberate’ *risuonare* ‘resound’, *rotolare* ‘rotate’, *salire* ‘rise’, *saltare* ‘jump’, *sbandare* ‘skid’, *scarseggiare* ‘become scarce’, *scendere* ‘descend’, *suonare* ‘sound’, *ticchettare* ‘tick’, *vivere* ‘live’
aanhoudden ‘continue’, *afenemen* ‘lower’, *behorend* ‘belong to’, *bloeien* ‘bloom’, *dalen* ‘descend’, *draaien* ‘rotate’, *duren* ‘last’, *helen* ‘heal’, *leven* ‘live’, *lopen* ‘walk/run’, *onderduiken* ‘sound’, *ontbreken* ‘fail’, *rijzen* ‘rise’, *rotten* ‘rot’, *ruilen* ‘change’, *schijnen* ‘shine’, *slippen* ‘skid’, *springen* ‘jump’, *struikelen* ‘stumble’, *tikken* ‘tick’, *uitkomen* ‘bud’, *vellen* ‘fell’, *verdrinken* ‘affogare’, *verrotten* ‘rot’, *weergalmen* ‘reverberate’, *weerklinken* ‘resound’, *zinken* ‘sink’, *zwemmen* ‘swim’

4.2.6. Weather Verbs

There is debate as to whether verbs that describe meteorological phenomena are unaccusative (Levin & Rappaport Hovav, 1995). Dutch weather verbs predominantly choose *hebben* in compound tenses, while in Italian there is free variation between *avere* and *essere.*

Table 4.6. List of *Weather Verbs* in Italian & Dutch

WEATHER VERBS
nevicare ‘snow’, *piovere* ‘rain’, *piovigginare* ‘drizzle’, *tremare* ‘quake’
donderen ‘thunder’, *motregen* ‘drizzle’, *regenen* ‘rain’, *sneeuwen* ‘snow’

4.3. Extracted Sentences

As I said before, I am considering auxiliary verb selection to be the surface level result of the unaccusative split; unaccusative verbs take BE and unergatives take HAVE. To produce the corpus of unaccusative and unergative verbs, sentences were extracted that contained either auxiliary with the past participle of a verb in one of the aforementioned categories.

4.3.1. Italian

Even with relatively free word order in Italian, auxiliary verbs occur immediately before the corresponding past participles, but may be separated by an adverb (Napolitano & Devine, 1979) (all examples in this section taken from the Europarl corpus):

(54) a. Il dittatore è caduto.
the dictator BE fell
'The dictator has fallen.'

b. La Commissione ha già affrontato questo punto...
the commission HAVE already addressed this point
'The Commission has already addressed this point...'

Sentences with this pattern are relatively easy to extract. I included a list of stopwords as to avoid extracting phrases with a nominal use of words that are also past participles, as below:

(55) Constato con soddisfazione che vi è un corso favorevole...
noticed with satisfaction REL to you BE a course acceptable
'I noticed with satisfaction that you have adopted an acceptable course...'

Corso, when used nominally, can mean 'course' or 'run', but it is also the past participle of the verb *correre* 'run'. The sentence extraction algorithm did not extract sentences like the above, but did extract sentences when *corso* is a past participle:

(56) Un elefante non ha mai corso più rapidamente di una zebra.
an elephant not HAVE ever run-PP more quickly of a zebra
‘An elephant has never run as fast as a zebra.’

4.3.2. Dutch

In Dutch, the word order is more rigid with auxiliary verbs coming in second position and participles coming at the end of the phrase (Donaldson, 1997). In the example below, when a complex NP (*de leiders van de demonstranten*) is the subject, the auxiliary (*hebben* is the sixth word in the sentence and the past participle (*overgegeven* is the fifteenth:

(57) De leiders van de demonstranten hebben zich na een massale
the leaders of the demonstrators HAVE themselves after a massive
inzet van het leger overgegeven.
deployment of the army surrendered
‘Following heavy military intervention, the leaders of the demonstrators have surrendered.’

In subordinate clauses, word order is different, with the auxiliary in final position:

(58) Ik hoop dat de rapporteur al deze voorstellen kan steunen als zij
I hope that the rapporteur all these proposals can-3S support all she
er nog een nachtje over geslapen heeft.
there once more night on slept HAVE
‘I hope that the rapporteur will be able to support all these proposals if she is allowed to sleep on it.’

Since there is no distance restriction between auxiliary verb and its participle, I extracted all sentences that contained an auxiliary and past participle of any of the verbs in the previously defined categories. I realize that this may unintentionally include some sentences that use participles in an adjectival manner, introducing some noise into the data set, but anticipate that similar results will ultimately be achieved.

4.4. Features and Experimental Setup

For each language, sentence-level feature values were extracted. I included 'bag-of-words' features, as well as binary values for membership in any of the various verb categories. For Italian, the presence of *ne* was included as a feature, since it is strongly coindicated with unaccusative verbs.

Due to the lack of consistent multilingual tools, semantic role labeling was done on the corresponding English translation of each sentence in the extracted sentences. Temporal and locative information was also automatically marked. As discussed, the subject of some unaccusative verbs are underlying patients, and there are no patients in unergative verbs. Levin and Rappaport Hovav's discussion of the importance of the internal arguments in unaccusative verbs also lend credence to the idea of developing metrics for *agenthood* and *patienthood.*

Since I am building a sentence-level classifier, an average score of agenthood and patienthood was calculated for each verb, and was compared to the average agenthood and patienthood as calculated for a substantial portion of the Europarl corpus, and was considered a binary feature. Similar average values were computed for temporal and location information, and again were used as binary features.

The first run, which I will consider the baseline, will include membership in the various verb class as features, the set of semantic role labeling features, and the *ne* feature (Italian only). For the second run, the tree-based features will be added. The third run will include the a "bag of words" featureset, and the fourth auxiliary verb choice.

4.5. Semantic Role Labels and PropBank

SENNA uses a Neural Network-based approach for part-of-speech tagging, chunking, named entity recognition, and semantic role labeling (Collobert & Weston, 2007). SENNA's semantic role labeling uses the PropBank (Palmer et al., 2005) convention of assigning up to six roles (ARG0-5) to each verb. These arguments depend on the *frame* of each verb, the manner in which it is used, and the number of embedded clauses. Below is a simple example from PropBank:

(59) ...[$_{Arg0}$ Sotheby's] ...offered [$_{Arg2}$ the Dorrance heirs] [$_{Arg1}$ a money-back guarantee]

In PropBank, ARG0 is considered to be an prototypical agent, and ARG1 is considered to be a prototypical theme or patient. Since many major treatments of unaccusativity and unergativity rely on the notions of agenthood and patienthood, I have only considered roles ARG0 and ARG1 in my study. ARG0 is expected to be the primary argument of unergative verbs, and ARG1 of unaccusatives. Below are simple example outputs from SENNA for an unaccusative verb and unergative verb respectively.

The only argument of the unaccusative *sink*, *the Samina*, is tagged as ARG1. The argument of the unergative *shout*, *someone*, is tagged as ARG0, as is the first argument of the transitive verb *hear*. Locational and temporal information is tagged where applicable.

4.6. Event Classification Approximation

With my feature selection, I aim to mimic some of the results of recent work in event classification. As we have seen, the telic/atelic distinction is extremely

Table 4.7. Sample Unaccusative Sentence in SENNA

"A few months ago, the Samina sank in the Aegean."					
A	DT	B-NP	O	-	B-AM-TMP
few	JJ	I-NP	O	-	I-AM-TMP
months	NNS	E-NP	O	-	I-AM-TMP
ago	RB	S-ADVP	O	-	E-AM-TMP
,	,	O	O	-	O
the	DT	B-NP	O	-	B-A1
Samina	NNP	E-NP	S-MISC	-	E-A1
sank	VBD	S-VP	O	sank	S-V
in	IN	S-PP	O	-	B-AM-LOC
the	DT	B-NP	O	-	I-AM-LOC
Aegean	NNP	E-NP	S-LOC	-	E-AM-LOC
.	.	O	O	-	O

Table 4.8. Sample Unergative Sentence in SENNA

"All I could hear was someone shouting down in the courtyard."						
All	DT	B-NP	O	-	O	O
I	PRP	E-NP	O	-	S-A0	O
could	MD	B-VP	O	-	S-AM-MOD	O
hear	VB	E-VP	O	hear	S-V	O
was	VBD	S-VP	O	-	O	O
someone	NN	S-NP	O	-	O	S-A0
shouting	VBG	S-VP	O	shouting	O	B-V
down	RP	S-ADVP	O	-	O	E-V
in	IN	S-PP	O	-	O	B-AM-LOC
the	DT	B-NP	O	-	O	I-AM-LOC
courtyard	NN	E-NP	O	-	O	E-AM-LOC
.	.	O	O	-	O	O

important in determining factors of unaccusativity, especially in Dutch (van Hout, 2004).

Using the Stanford Parser (Klein & Manning, 2003), I parsed the extracted sentences to make use of certain elements of phrase structure. The program is not designed for Italian and Dutch, but it is certainly sufficient to analyze the English translations of the extracted sentences. Italian and Dutch both exhibit corresponding to- and for- style PPs, so the English translations should be a fair approximation.

While I was not able to use the programs described in the following sections, I did have access to portions of their code in order to approximate some of their features.

4.6.1. Lenci & Zarcone

In their 2008 paper, Lenci and Zarcone trained both event-type classifiers for Italian-language data. In order to do so, they have further classified Vendler's categories as explained in Chapter 2.

Using combinations of tenses and temporal information, my feature set mimics Lenci and Zarcone's *Aktionsart* classification. Not all other features from Lenci and Zarcone were mimicked in my study; as they themselves reported, adverbial features appear too infrequently to have a great impact on results.

With the parses from the Stanford Parser, I was able to consider as features the presence of various prepositional adjuncts which influence unaccusativity. As mentioned in the discussion of Chierchia's and Pustejovsky's proposals, to- and for-phrases are particularly important in determining the headedness and event structure of verbs that demonstrate split intransitivity.

As I noted previously, I only extracted verbs appearing in periphrastic constructions, as I have considered this the most reliable surface level feature for unaccusativity. In both Italian, and Dutch, present tense constructions do not contain the syntactic variability that is seen in the simple past. Lenci & Zarcone's classifier accuracy rates were on par with hand-annotated results by trained linguists, offering further proof for the inherent difficulties with this type of analysis (Zarcone & Lenci, 2008). Even linguists do not always agree on event-types.

4.6.2. Im & Pustejovsky

Im and Pustejovsky have adopted Lenci and Zarcone's work for English verb event-structure (Im & Pustejovsky, 2010). Again, I had access to portions of this code, so I was able to imitate some of the feature selection. The goals of this paper and my work have a different scope, but we are certainly investigating similar phenomena. As I described above, I attempted to incorporate some event type information in my feature selection for the unaccusativity classifier. Im and Pustejovsky are building an entire lexicon of event-based implicatures, and I believe that a more comprehensive treatment on the event-structure of unaccusative verbs will prove to be a useful experiment.

4.7. Prepositional Phrases

In order to approximate telicity, I used categories of prepositional phrases, as outlined in Im & Pustejovsky's event classifier (2010). Each of these categories of PPs is a binary feature in my classifier. As discussed in Van Valin, and Chierchia above, Dowty devised tests to determine membership in Vendler's event categories (Van Valin Jr, 1990). To reiterate, for-PPs are allowed with atelic phrases (accomplishments and achievements), in-PPs with telic phrases (states and activities).

As we have seen, event types can change depending on other linguistic context. *Run* is normally an activity, and therefore atelic, but when combined with a for-PP, it becomes a telic event (Chierchia, 2004). Of course, even prepositions can be ambiguous, *run for an hour* vs. *run for your health.*

4.7.1. Change of Location Prepositions

This group contains *to*, and therefore will represent to-PPs as a whole. They represent changes in location, and therefore directed motion. Manner of motion verbs in Italian and Dutch usually take HAVE, but with a change of location, they take BE (Levin & Rappaport Hovav, 1995).

Table 4.9. Change of Location Prepositions

C.O.L. PREPOSITIONS
from, to, into, onto, out_of, across, close_to, aboard, past, away, out, on, off

4.7.2. Locative Prepositions

As noted, in-PPs are a test for telicity and sound strange with atelic events (Chierchia, 2004).

Table 4.10. Locative Prepositions

LOCATIVE PREPOSITIONS
in, at, on, under, above, over, below, next_to, close_to, near, beyond, beneath, behind, underneath, upon

4.7.3. Path Prepositions

This group also provides a directionality to manner of motion verbs.

Table 4.11. Path Prepositions

PATH PREPOSITIONS
along, by, around, round, via

4.7.4. Directional Prepositions

Directional Prepositions add a directional interpretation to manner of motion verbs. However, *for* has many uses. Perhaps this is an unfortunate name for this category of PPs, considering the importance of directed motion, but the atelic use of *for* is likely more common (e.g., *run for three hours* vs. *run for the door*). I will refer to this group as for-PPs.

Table 4.12. Directional Prepositions

DIRECTIONAL PREPOSITIONS
towards, toward, for, down, up

4.7.5. Neutral Prepositions

These PPs have not been ascribed any significance in matters of split intransitivity.

Table 4.13. Neutral Prepositions

NEUTRAL PREPOSITIONS
about, absent, after, against, along, alongside, amid, amidst, among, amongst, as, aside, astride, at, athwart, atop, barring, before, below, beside, besides, between, betwixt, but, circa, concerning, despite, during, except, excluding, failing, following, given, including, inside, like, mid, minus, next, notwithstanding, of, opposite, outside, pace, per, plus, pro, qua, regarding, save, since, than, through, thru, throughout, thruout, till, times, unlike, until, versus, vs, vice, with, within, without, worth]

4.8. Methodological Issues

My experiment as designed has some shortcomings, but due to the corpus size, I believe I did not introduce too much noise into the data set. Relying on English translations of the Italian and Dutch unaccusative and unergative sentences may skew the counts of ‘agenthood’ and ‘patienthood’ somewhat, especially with Italian verbs like *piacere* ‘please’ which is commonly translated into English as *like*; the linear order of agent and patient roles are switched, but in both languages, there is still one of each, so raw counts would not be effected. Presence of a temporal phrase does not always necessarily modify the event type encoded in the verb. It may merely ground the event in time.

4.8.1. Improvements and Future Work

Ideally, I would have been able to annotate corpora in Italian and Dutch with event types and internal vs. external arguments, or use extant tools for event classification. Unfortunately, this was beyond the scope of the project, and in the case of previous English and Italian *Aktionsart* classification work, they were not in a polished enough state for me to apply them to my study. To my knowledge, there is not a comparable Dutch resource available.

Another approach would have been to build a verb sense-level classifier, instead of a sentence-level classifier. Again, this would require more intense annotation work and lemmatization. In this scenario, we could use instances of impersonal passivization, resultative phrases, and other diagnostics as features for each verb lemma, instead of relying on approximations of unaccusativity on the sentence level. The extraction algorithm only looked at the verb and its auxiliary, so it is probable that some transitive alternants were extracted as well.

4.9. Hypotheses

Based on the various semantic and syntactic accounts presented in the previous Chapter, as well as the various diagnostics outlined in Chapter 1, I believe that unaccusativity will be testable in the computational linguistic domain, and will frame my predictions in terms of their predictive power in regards to unaccusativity.

To begin with the clearer cases, I expect *ne* to appear almost exclusively with Italian *essere* sentences. the sentences will be selected on the basis of their auxiliary verb, so I expect it to be a very strong indicator of unaccusativity; so strong, in fact, that I will perform the bulk of my trials without it because I expect that it may skew my results. For the semantic role label-based features, I also expect low agenthood and high patienthood to be strongly associated with unaccusativity.

I expect that BE sentences will have more locative phrases, since unergative manner of motion verbs may be unmarked, and they pair with BE with the addition of directional information. Particularly in Dutch, I expect theoretically telic BE verbs to contain less temporal adjuncts.

I expect telic event types (accomplishments and achievements) to be strongly coindicated with unaccusativity; the larger concern will be how to interpret the features as to make such a judgment possible. Locative (in-PPs) and COL (to-PPs) features should be coindicated with unaccusativity as well, as they ground the right-headed resulting state of achievements and accomplishments. For-PPs provide temporal information, so I expect them to pair with HAVE sentences.

CHAPTER 5

Discussion of Results

5.1. Results

MaxEnt and Naïve Bayes classifiers were run four times each for both languages, with additional features being added for each trial. For Italian, 2351 BE sentences were extracted, and 1560 HAVE. In Dutch, there were 4168 Dutch BE sentences and 2756 with HAVE.

Due to the relatively small size of my data set, the sentences were shuffled for a 5-wise cross-validation of training and test sets.

Table 5.1. Classifier Accuracy Scores

Classifier Accuracy Scores				
Trial	IT(NB)	NL(NB)	IT(ME)	NL(ME)
1. Verb Class & SRL	60.4%	60.4%	57.2%	70.4%
2. Features from 1 + Tree	58.5%	59.9%	57.2%	73.6%
3. Features from 2 + BOW	66.4%	60.9%	66.2%	69.6%
4. Features from 3 + Aux.	74.7%	66.7%	74.6%	72.9%

5.2. The Trials

5.2.1. Trial 1 - Verb Class, Semantic Roles, Ne

The first run included the various verb class features, semantic role labeling information, as well as *ne.* Although *ne* is strongly linked to unaccusativity in Italian, there were only 57 instances in the extracted sentences (1.45%), so it was incldued in the baseline run. There is no analagous feature in Dutch, so the baseline just included the verb class and semantic role information.

In Italian, the accuracy of this run was 60.4% (I will cite accuracy scores from the Naïve Bayes classifier). As we might expect, nearly 34% of HAVE verbs exhbited an above-average agenthood, with only 26% of BEs. 37.6% of the BE verb instances had an above-average patienthood (ARG1), however over 60% of HAVEs did as well, suggesting that the agenthood/patienthood distinction is muddied considerably by the fact that the HAVE sentences contained transitive alternants, inherently containing an ARG1 role.

Approximately 66% of HAVE sentences contained a locative expression, as opposed to 41% of unaccusatives, at odds with the hypothesis that changes in location are linked with unnacusatives. The change in location PP (frequently with *to*) turns processes into transitions, thereby grounding the right-headed event of the resutling state of the action. However, SENNA may mark any locative expression, so marked locatives may merely be providing background information. As we would expect, unaccusatives showed fewer temporal expressions than HAVE verbs in Italian, in line with the telic/atelic distinction.

I will not go into every value for each featureset, only highlighting differences when they are particularly useful or unexpected.

Table 5.2. Results of Italian Trial 1

Italian Trial 1		
	BE	HAVE
A0 above avg.	26.6%	33.6%
A1 above avg.	37.5%	60.7%
LOC above avg.	40.9%	66.5%
TMP above avg.	31.3%	48.5%

In Dutch, the difference in agency is even more striking. Approximately 46% of HAVE verbs have an above-average agenthood, compared to 27% of unaccusatives. Interestingly, Dutch unaccuastives exhibited more locative expressions than their

HAVE counterparts (56% to 38%), suggesting that the temporal information may not be irrelevant to (or at least is coindicated with) Dutch unaccusativity. The remaining features show a similar distribution to Italian, again suggesting that semantically marked temporal expressions may not be a useful feature unless they specifically modify the appropriate verb, although the discrepancy is less severe in Dutch.

Dutch BE verbs showed fewer temporal expressions than HAVE verbs, again in line with the telic/atelic distinction. Perhaps it is unwise to compare Dutch results to Italian results, but Dutch BE verbs had even fewer temporal expressions than the Italian BE verbs, further suggesting the singularity of telicity to Dutch unaccusativity.

Table 5.3. Results of Dutch Trial 1

Dutch Trial 1		
	BE	HAVE
A0 above avg.	27.8%	46.3%
A1 above avg.	31.7%	34.4%
LOC above avg.	56.2%	43.2%
TMP above avg.	26.3%	27.4%

However, sparsity of features will prove to be a problem, as among the most informative features according to NLTK's built-in method of the same name, were sentences without agents and patients in Italian and being a member of the *body* verb class and having a low score for locatives in Dutch. Verb class did not play a major role throughout the study, but in Dutch 88% of the 25 extracted*body* verbs patterned with BE. This is the case of a very sparse featureset having an outsized impact on the data.

In both languages, the majority of extracted verbs came from the predefined 'known UA/UE alternants' class. *Roll*, *run*, and *sound* verbs numbered in the several hundreds for each language, with a handful of *weather* and *body* verbs.

5.2.2. Trial 2 - Tree features

Trial 2 included all of the tree-based features as outlined above, however accuracy scores for both Italian and Dutch were slightly lower than in the previous trial, suggesting that these features introduced more confusion into the data.

In Dutch, for-PPs (indicators of telicity) occurred in 28% of HAVE verbs, and only 12% of BE verbs, indicating a strong cooccurence of atelic verbs with unaccusatives, as we expect. In Italian, wehere telicity is not the sole indicator of unaccusativity, but is strongly linked (Chierchia, 2004), there is also a greater preponderance of for-PPs with HAVE verbs (16%) as with BE verbs (10%).

In-PPs behaved differently across languages; in Dutch they were more strongly coindicated with BE verbs, in Italian they were more strongly coindicated with HAVE verbs. This could be for a number of reasons - perhaps the locative preposition category was too broadly defined, or it could simply be due to the variability of the preposition *in* as seen in the examples from Chierchia:

(45d) Gianni ha corso in giardino in un minuto.
Gianni BE run in garden in a minute.
'Gianni ran in the garden in a minute.'

Due to the nature of the featureset, had they appeared in the corpus, both sentences would be given the in-PP feature. Contrary to our hypotheses, Italian

HAVE verbs cooccur more frequently with *change of location* (COL) PPs; contrastively, in Dutch, where telicity is argued to be the main factor in determining unaccusativity, the COL features occurs more frequently with BE verbs. While not an unexpected result, since telicity and COL features are frequently found together in unaccusatives, I expected the correlation to be weaker in Dutch.

Table 5.4. Tree Features

%age of Verbs with various Tree Features				
Feature	BE-IT	HAVE-IT	BE-NL	HAVE-NL
COL.	50.9%	88.6%	75.7%	50.1%
DIR	9.8%	16.2%	12.2%	28.1%
LOC	40.9%	66.5%	56.2%	37.7%
PATH	5.9%	10.0%	9.1%	17.5%

While some of the PP-based features fit into our theoretical paradigm, the classifier performed worse on this trial. A possible cause is the NEUTRAL PP feature. This feature occurred in the majority of sentences, so it is probable that it muddied the waters, diminishing the impact of the other features. There was no clear indication of neutral PPs appearing more with BE or HAVE verbs.

In Italian and Dutch, semantic features original to Trial 1 remained the most informative; according to the most informative features method, a low locative number and high agenthood being more influential than various PP features. As discussed above, the ambiguity of certain PPs likely contributed to this trial performing worse.

5.2.3. Trial 3 - Bag-of-Words

The addition of bag-of-words features increases the accuracy of the classifier, making up for the loss when the tree-based features were added. The fact that the bag-of-words had such a relatively a large influence could mean that our core feature space does not contain enough features, or that their predictive power is not robust enough.

5.2.4. Trial 4 - Auxiliary Verbs

With the addition of the various auxiliary verbs as features, the classifier performance again improves. Naturally, this was expected, as auxiliary verb choice was considered the prime surface-level diagnostic for unaccusativity. Indeed, the auxiliary BE was considered among the most influential for both languages. However, it is not a perfect indicator, as all of the auxiliaries occur frequently in other roles.

Ultimately, the performance of the classifier is respectable, but not perfect. Scores of nearly 75% for Italian and 67% for Dutch are certainly better than chance, but leave something to be desired.

5.3. Potential Improvements to Features

As discussed in Chapter 4, there were some methodological concerns that may have impacted the quality of the data. The respectable classifier performance may very well be improved if it is trained on hand-selected examples of VPs containing verbs with unaccusative/unergative alternants. A certain degree of noise was introduced by relying on the English translations for semantic role and syntactic parses in an attempt to build a language-independent unaccusative classifier. The notion of unaccusativity varies widely cross-linguistically, and even intra-linguistically, so future trials would benefit from incorporating language-specific NLP tools in an attempt to identify components of unaccusativity on a per-language basis.

CHAPTER 6

Conclusions

Levin & Rappaport conclude that unaccusativity is 'syntactically encoded and semantically determined' (1995). In my attempt to frame the investigation around individual verbs that exhibit split intransitivity, I set out to quantify the internal and external features (semantic) that allow verbs to appear in (syntactic) constructions with either auxiliary, and believe that I have achieved this to a degree. From a theoretical perspective, unaccusativity is a wide-ranging phenomenon, with multiple manifestations within individual languages, language families, and across languages. The notion raises fundamental questions of the interaction between syntax and semantics, with clear syntactic manifestations of various semantic roles, event structure, and temporal relations. The understanding of the phenomenon has advanced from early descriptions of underlying objects and patients to more sophisticated accounts dependent on event structure and notions of telicity. Attempts to categorize and classify groups of unaccusative verbs into categories are imprecise and break down cross-linguistically, as do descriptive explanations of auxiliary verb selections.

One can wax philosophical on the notions of agenthood and our control over events in the universe , but it has proven to be difficult to quantify these notions.

6.1. What have we learned?

Notions of agents and patients are imperfect, and semantic role labelers are imperfect tools, so even though the results indicate a strong coincidence of unaccusative verbs and lack of an agent, we cannot say that all unaccusative verbs in Italian or Dutch lack an agent. The motion verbs, which we have discussed throughout this thesis, disprove this notion out of hand. The distinction is nebulous, but a strong coindication is suggested.

I believe that a more precise event-classification framework, based on interactions between adjunctive PPs and adverbials would improve my results. Better defined categories of PPs would more clearly delineate event classes, and due to ambiguity of PPs, the effort would be greatly helped by hand annotating data before building a classifier. I leave the application of event classifiers to unaccusative verbs for future research, as I believe temporality in unaccusativity is a small subset of larger interactions of Aktionsart.

Many other features of unaccusativity, such as locative inversion and resultative phrases would require an annotation effort before they could be considered useful features in this regard.

6.2. Is this really a useful class?

While many of my results suggest that unaccusativity can be quantified to some degree, ultimately the ambiguity in the results suggest that the notion of unaccusativity is not a useful class. Undoubtedly, both syntax and semantics are at play here; however this in itself does not make it interesting. Unaccusativity is the name that has been given to a range of phenomena that exhibit a complex interaction between agency, notions of time, verbal aspect, and historical quirks.

I have attempted to examine two languages that show a relatively high degree of unaccusativity, but even the millions of sentences of Europarl data contained a few thousand instances of these variable behavior verbs. Ultimately this is a 'chicken and egg' question; is auxiliary choice a result of initial unaccusativity (syntax), or do the semantics of a verb determine whether or not a verb is used unaccusatively (i.e., auxiliary choice selection)?

6.3. To have or not to have

Chierchia has proposed that the sole factor in (Italian) auxiliary verb selection is *subject-affectedness*, as *essere* goes with passives, reflexives, unaccusatives, and impersonals (Chierchia, 2004), all of which are subject to type-shifting operations. Benveniste (1966), in his treatise on the functions of *to be* and *to have* describes the two thusly:

> *Être* establishes an intrinsic relationship of equivalence between the two terms which it joins: it is the consubstantial state. In contrast, the two terms joined by *avoir* remain distinct; the relationship between them is extrinsic.

Perhaps the question of unaccusative verbs and auxiliary selection comes down to control. BE usages are generally telic, have no agent, and manifest when their action is modified by a change of location; HAVE usages are generally atelic and have an agent who has some say in the matter. The boundary between the two is different in different languages, indicating that there is no unique characteristic the separate the two cross-linguistically. I hope that this study has provided some insight into the unaccusativity phenomenon by means of utilizing modern techniques that have not been previously applied. To be or not to be is an interesting problem to have.

References

Alexiadou, A., Anagnostopoulou, E., & Everaert, M. (2004). *The unaccusativity puzzle.* Oxford University Press.

Benveniste, E. (1966). *Problèmes de linguistique générale* (M. Meek, Trans.). Gallimard.

Burzio, L. (1986). *Italian syntax: A government-binding approach.* Springer.

Calchini, E. e. a. (2011). *Dante search: Commedia edizione elettronica lemmatizzata.* Retrieved from `http://dante.di.unipi.it:8080/DanteWeb/`

Centineo, G. (1996). A lexical theory of auxiliary selection in italian. *Probus*, *8*, 223–271.

Chierchia, G. (2004). A semantics for unaccusatives. In M. E. A. Alexiadou E. Anagostopoulou (Ed.), *The unaccusativity puzzle.* Oxford University Press.

Collobert, R., & Weston, J. (2007). Fast semantic extraction using a novel neural network architecture. In *Annual meeting-association for computational linguistics* (Vol. 45, p. 560).

D'Alessandro, R. (2001). On impersonal si constructions in italian. In *Console x proceedings: Proceedings of the tenth conference of the student organization of linguistics in europe* (pp. 1–15).

Donaldson, B. (1997). *Dutch: A comprehensive grammar.* Routledge.

Dowty, D. (1979). *Word meaning and montague grammar: The semantics of verbs and times in generative semantics and in montague's ptq* (Vol. 7). Springer.

Im, S., & Pustejovsky, J. (2010). Annotating lexically entailed subevents for textual inference tasks. In *Twenty-third international flairs conference.*

Klein, D., & Manning, C. (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st annual meeting on association for computational linguistics-volume 1* (pp. 423–430).

Koehn, P. (2005). *Europarl: A parallel corpus for statistical machine translation.*

Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity: At the syntax-lexical semantics interface* (Vol. 26). The MIT Press.

Loper, E., & Bird, S. (2002). Nltk: the natural language toolkit. In *Proceedings of the acl-02 workshop on effective tools and methodologies for teaching natural language processing and computational linguistics - volume 1* (pp. 63–70). Stroudsburg, PA, USA: Association for Computational Linguistics.

Maiden, M. (1995). *A linguistic history of italian.* Longman.

Moretti, G., & Orvieto, G. (1979). *Grammatica italiana: Volume 1 il verbo.* Editrice Benucci.

Napolitano, A., & Devine, M. (1979). *Manuale di grammatica italiana.* Anma Libri.

Palmer, M., Gildea, D., & Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, *31*(1), 71-105.

Perlmutter, D. (1978). Impersonal passives and the unaccusative hypothesis. In J. J. Jacobs et al. (Eds.), *Proceedings from the fourth meeting of the berkeley linguistic society* (pp. 157–189).

Perlmutter, D. (1989). Multiattachment and the unaccusative hypothesis: the perfect auxiliary in italian. *Probus*, *1*(1), 63–120.

Priebsch, R. (1948). *The german language.* Faber & Faber.

Pustejovsky, J. (1995). *The generative lexicon.* MIT press.

Pustejovsky, J., & Busa, F. (1995). Unaccusativity and event composition. *Temporal Reference, Aspect and Actionality*, *1*, 159–177.

Randall, e. a., J. (2004). Acquiring unaccusativity: A cross-linguistic look. In M. E. A. Alexiadou E. Anagostopoulou (Ed.), *The unaccusativity puzzle.* Oxford University Press.

Shetreet, E., Friedmann, N., & Hadar, U. (2010). The neural correlates of linguistic distinctions: unaccusative and unergative verbs. *Journal of Cognitive Neuroscience*, *22*(10), 2306–2315.

Sorace, A. (1993). Unaccusativity and auxiliary choice in non-native grammars of italian and french: asymmetries and predictable indeterminacy. *Journal of French Language Studies*, *3*(01), 71–93.

Sorace, A. (2004). Gradience at the lexicon-syntax interface. In M. E. A. Alexiadou E. Anagostopoulou (Ed.), *The unaccusativity puzzle.* Oxford University Press.

Tortora, C. (2001). Evidence for a null locative in italian. *G. Cinque, G., Salvi (eds.), Current studies in Italian syntax, North-Holland: Amsterdam*, 313–326.

van Hout, A. (2004). Unaccusativity as telicity checking. In M. E. A. Alexiadou E. Anagostopoulou (Ed.), *The unaccusativity puzzle.* Oxford University Press.

Van Valin Jr, R. (1990). Semantic parameters of split intransitivity. *Language*, 221–260.

Zaenen, A. (1988). *Unaccusative verbs in dutch and the syntax-semantics interface.* CSLI.

Zarcone, A., & Lenci, A. (2008). Computational models of event type classification in context. In *Proceedings of lrec* (pp. 1232–1238).