

CHAPTER 3. A QUESTIONNAIRE STUDY FOR CATEGORIZING VERB PHRASES ACCORDING TO THEIR EVENT TYPE

3.1 Introduction

This chapter presents a questionnaire study that was designed for the purpose of eliciting the judgments and interpretations needed to categorize a verb phrase as telic or atelic.¹⁰ In the past, syntacticians and semanticists have determined the telicity of verb phrases by eliciting these sorts of judgments and interpretations from themselves and perhaps a few of their colleagues. The participants in this questionnaire study are untrained native speakers of English. This kind of test is reminiscent of “norming” tests commonly used in the area of psycholinguistics. However, norming usually refers to collecting information on whether a lexical item is biased towards a particular kind of usage. For example, verbs may be normed in order to determine whether native speakers more often use them transitively or intransitively. It is important to note that in this dissertation, telicity is considered to be a categorical property of a verb phrase, rather than a matter of bias.¹¹ The primary purpose of the questionnaire is to work towards the development of an objective measure of telicity, so that one does not have to have extensive training in order to determine the telicity of an event.

¹⁰ I recently learned that a graduate student named Andrea Proctor at Northwestern University is concurrently running a questionnaire study that is similar to the one presented in this chapter (Mike Dickey, Andrea Proctor, personal communication). I have not yet seen a report of the results from her study.

¹¹ An atelic verb may be used in an atelic verb phrase or a telic verb phrase. This is due to the compositional nature of telicity, which is discussed in detail in chapter 2.

This questionnaire study is my first attempt to collect responses for the purpose of telicity categorization, and the study is still in progress. The results presented in this chapter are preliminary, and they indicate that the questionnaire itself needs adjustment. In the short term, this questionnaire helps to clarify which tests are most useful for making telicity categorizations and which kinds of verbs the tests work best for. For example, although achievements and accomplishments are both telic events, the responses to a particular question for achievement events seem to differ systematically from the responses for accomplishments. The reasons for expecting different questionnaire results for Vendler's (1957, 1967) event types are discussed briefly in the introduction of this chapter and in more detail in chapter 2. In future work, I hope that the responses to an improved version of this questionnaire will be helpful for interpreting the results of the psycholinguistic experiments in this dissertation and elsewhere. For example, the questionnaire results may indicate previously overlooked differences between verbs within a category such as "obligatorily transitive telic."

Three classic telicity tests are included in the questionnaire study. Here I will describe the three question types and explain my predictions for each of them.

3.1.1 Question Type 1: "For an hour" Versus "In an hour" Forced Choice

In this type of question, the participant is presented with two similar sentences and is asked to choose which of the two is most natural. One of the sentences ends with *for* <some time interval>, and the other ends with *in* <some time interval>. Two examples are given in (85) and (86).

(85) *Which of the two sentences is most natural?*

Trinity ran a mile for five minutes.

Trinity ran a mile in five minutes.

(86) *Which of the two sentences is most natural?*

Trinity ran for five minutes.

Trinity ran in five minutes.

The event *ran a mile* is a canonical example of a telic accomplishment. Many philosophers and linguists, starting as early as Vendler (1957, 1967) and Kenny (1963), have claimed that such events are congruent with the adverbial modifier *in an hour* and not with *for an hour*. (The time interval should not be relevant; it can be adjusted to a plausible duration.) The predicted answer in (85) is *Trinity ran a mile in five minutes*. In contrast with the event *ran a mile*, the event *ran* is a canonical example of an atelic activity. Activities are considered to be congruent with the adverbial modifier *for an hour* and not with *in an hour*. Thus, the predicted answer in (86) is *Trinity ran for five minutes*.

For most events, aspectual coercion is expected to affect the results of the *in an hour* versus *for an hour* test. For example, it is acceptable to use the telic event *punched the agent* with the durative adverbial in *Neo punched the agent for five minutes*, because the aspectual coercion operation yields an iterative interpretation for *punch the agent*. Aspectual coercion is discussed at length in chapter 2. The important point for this chapter is that I do not expect that *for an hour* will be chosen for atelic events 100 percent of the time or that *in an hour* will always be chosen for telic events. This could only happen if the participant always chose the sentences that do not require coercion. Rather, I predict that the *for an hour* adverbial will be chosen more often for atelic events than

for telic events and that the *in an hour* adverbial will be chosen more often for telic events than for atelic events.

3.1.2 Question Type 2: The Imperfective Paradox Test Passages with Questions

The imperfective paradox is even older than the *in an hour* versus *for an hour* test; it was discussed by Aristotle, then later by Ryle (1954), Vendler (1957, 1967), and Kenny (1963). The imperfective paradox is that the imperfective form of a sentence (like *building a ship*) can be true even if the perfective form (*built a ship*) will never become true. This state of affairs occurs when an accomplishment is interrupted in the middle of the event. If Morpheus is building a ship and he quits before the ship comes into existence, *Morpheus built a ship* will not become true. In contrast, the perfective form of an atelic event becomes true almost immediately after the imperfective form is true. If Naiobe is dancing is true, no matter when she quits, *Naiobe danced* or *Naiobe has danced* will be true.

In the questionnaire study, the imperfective paradox is used as a test in the form of brief passages followed by questions, such as the examples in (87) and (88).

(87) *In the school cafeteria, a boy was eating a cupcake. While he was eating the cupcake, a food fight started. He stopped eating the cupcake and threw his asparagus at the girls at the next table.*

Question: Did the boy eat the cupcake?

(88) *At a fiesta on Saturday, a girl was dancing. While she was dancing, a fight broke out. She had to stop dancing, so that she could get out of the way.*

Question: Did the girl dance?

In (87), *eating a cupcake* is the event being tested. In the event structure literature, an event like *eating a cupcake* is considered a telic accomplishment (Tenny, 1992). The prediction for *Did the boy eat the cupcake?* is that the answer should be “no” more often than “yes”, because he didn’t finish eating the cupcake. In (88), the activity *dancing* is a classic atelic event. Thus, the prediction for *Did the girl dance?* is that the answer should be “yes” more often than “no”.

If the predictions turn out to be correct for the classic examples of telic and atelic verbs, then the test should be a useful diagnostic for determining the telicity of other events. However, the predictions are less clear for telic achievements and atelic states. Achievements such as *noticing the painting* are also a bit confusing in the imperfective paradox test. If Neo is noticing the painting, and then he stops noticing the painting in the middle, did he notice the painting? This is difficult to answer because noticing a painting, like other achievement verbs, is instantaneous.

The imperfective paradox is an effective test for some states but not for others. The state *being happy* sounds natural enough in the imperfective paradox test. If Trinity is happy, no matter when she stops being happy, it will be true that she was happy. This result is as predicted for an atelic event. However, consider the state *knowing Kung Fu*. This event will be tested by saying that Neo is knowing Kung Fu, and then he stops knowing Kung Fu. It will be true that he knew Kung Fu, as would be predicted for an

atelic event, but the test sounds very odd. It is unnatural to stop knowing Kung Fu, because knowing Kung Fu is an individual level predicate. A discussion of stage level versus individual level predicates is beyond the scope of this dissertation. I refer the reader to Carlson (1977) for more information on the distinction. What is important for the current discussion is that individual level predicates are generic, possibly permanent properties that an individual has. For that reason, it sounds unnatural to say that an individual level predicate *stops*.

3.1.3 Question Type 3: "It took an hour to" Judgments

Many philosophers and linguists have noted that telic events, but not atelic events, are acceptable in a sentence beginning *It took <some time interval> to* (Vendler, 1957, 1967; Dowty, 1979). For example, because *build the ship* is a telic event, (89)a is very natural. In contrast, *build ships* is atelic, so (89)b is unnatural.

(89)

- a. *It took a month to build the ship.*

- b. *It took a month to build ships.*

This test works well for most states (atelic) and achievements (telic). The sentence in (90) is unnatural, as it should be, since states are atelic.

(90) # *It took five minutes to be happy.*

The individual level state in (91) ought to be equally unnatural, since it is atelic. Some speakers may interpret the sentence to mean “It took five minutes to *begin* knowing Kung Fu.”

(91) # *It took five minutes to know Kung Fu.*

The sentence in (92) should sound fairly natural, since *notice the painting* is a telic achievement.

(92) *It took a minute to notice the painting.*

However, the interpretation of the event in (92) differs from the interpretation of the event in (89)a. Because *build a ship* is an accomplishment, the event is taking place throughout the time interval. Because *noticing the painting* is an achievement, the event is not taking place throughout the time interval. Rather, the noticing of the painting only occurs at the end of the time interval.

In the questionnaire, participants rate sentences like the ones in (89) through (92) according to their naturalness on a scale from one to seven. My prediction is that accomplishments will be rated as the most natural and that achievements will be rated second most natural. Activities should be rated more unnatural than natural. States are not included in this initial version of the questionnaire, because they sound so unnatural in the tests.

3.2 Methods

3.2.1 Participants

The questionnaire participants are all monolingual speakers of English. At the time of this report, there have been 24 participants. Since there are six presentation lists, four participants are included in each list. The participants were all undergraduate students at the University of Arizona taking an introductory course in linguistics. For their participation, students received additional points added to their lowest exam score. Since handedness has been shown to be a significant factor affecting performance on language-related tasks (Treadwell, Nicholas, and Bever, in preparation), the results from left-handed and right-handed participants are calculated separately. At the time of this report, all of the monolingual English-speaking participants have been right-handed (coincidentally).

3.2.2 Procedure

The questionnaire has three sections. In all of the sections, the questions are presented on a computer screen by the software DMDX.¹² The participants respond by pressing particular keys on the keyboard. Each of the three sections presents a different type of question. The participant completes four practice trials of each of the three types of questions at the beginning of the session while the experimenter is present. After the

¹² DMDX experimental software was programmed by Jonathan Forster at the University of Arizona in Tucson.

participant completes the practice trials for each of the three question types, the experimenter answers any questions that the participant has about the instructions.

The first third of the experiment is the *for an hour* versus *in an hour* forced choice section. On each trial in this section, the participant is presented with two similar sentences and is asked to choose which of the two is most natural. One of the sentences ends with *for <some time interval>*, and the other ends with *in <some time interval>*. An example trial is shown in (93).

(93) *Which of the two sentences is most natural?*

Mary arrived for an hour.

Mary arrived in an hour.

In order for the text to fit on the screen, the two sentences are presented with one on top and the other on bottom. They are asked to press the “up” arrow to select the top sentence or the “down” arrow to select the bottom sentence. The location of *for <some time interval>* and *in <some time interval>* was randomized in the original list. Half of the participants receive this original list, and the other half of the participants receive the exact opposite (*for <some time interval>* is on top in the opposite list when it is on the bottom in the original list, and vice versa). The forced choice section of the experiment was presented to the participants first so that the second and third test types would not influence the participants' choices.

The second third of the experiment is the imperfective paradox test reading passage section. On each trial in this section, the participant is asked to read a short passage and then to answer three comprehension questions about it. One of the

comprehension questions is relevant for event classification; the other two are distracter questions. An example trial is shown in (94).

(94) *At a roller skating rink last weekend, a girl was doing the hokey pokey. While she was doing the hokey pokey, her best friend called out to her. She stopped doing the hokey pokey and went to eat pizza with her best friend.*

Question 1: Did the girl go to the roller skating rink?

Question 2: Did the girl do the hokey pokey?

Question 3: Did the girl ignore her best friend?

All of the passages in this section, with the exception of the practice trials, had the following form: <Introductory context>, <subject noun phrase> <event in present progressive form>. <While/In the middle of> <pronoun referring to subject noun phrase> <event in present progressive form>, <some interruption occurred>. <pronoun referring to subject noun phrase> <stopped/had to stop> <event in present progressive form> <reason for stopping>. The experimental question always had the following form: <Did> <subject noun phrase> <event in bare form>?

The distracter questions were always questions that should be unambiguously answered “yes” or “no”. One of them should be answered “yes”, and the other should be answered “no”. The order of the three questions was randomized for each experimental item but held constant across experimental conditions (the four verb types). The reason for this was to avoid a pattern in having a “yes” or “no” response preceding the experimental question.

sentences, the events tested consist of the verb followed by the initial noun phrase. For example, the reduced relative clause sentence in (96) is tested as shown in (97).

(96) *The actress spotted by the writer left in a hurry.*

(97) *Please rate the naturalness of the following sentence...*
It took five minutes to spot the actress.

The reason for including the events from the maze experiment is that this will allow the questionnaire results to eventually be compared with the event categorizations that were already made for the maze experiment. If the questionnaire results differ from the event categorizations for the maze experiment, it will be interesting to see whether new categorizations based on the questionnaire results shed additional light on the maze experiment data.

The other twelve events include some canonical examples of activities, achievements, and accomplishments, as well as a few other events of interest. The canonical activities include *dance* and *push the cart*. The canonical achievements include *cross the finish line* and *notice the painting*. The canonical accomplishments include *drink the beer* and *eat the cupcake*. The non-delimited versions of these accomplishments, *drink beer* and *eat cupcakes*, were also included, since the unbounded nature of the theme makes them activities. The verb phrase *hit the tennis ball* was included because it describes an instantaneous event that is easily coerced into an activity; this kind of event is sometimes referred to as a “semelfactive” (Smith, 1991:55-

57). The event *stroke the cat* was also included because it may be a semelfactive. *Dance the Macarena* and *do the hokey pokey* were selected because they are dances that seem to have a delimited time course; they may be accomplishments.

Much care has been put into determining the time intervals used in the tests. The time intervals vary depending on the event. Ideally, the time interval should be one that is a very plausible amount of time for the event to take place, so that the time interval does not affect the participants' judgments on the questions. For each particular event, the time interval used is the same in both of the question types that use time intervals. For example, if *for an hour* is used in the first section with a particular event, then *It took an hour* will be used in the third section. An attempt was made to norm the time intervals, based on Zwaan (1996), but the methods used in that paper turned out not to work for this experiment. Instead, three judges carefully reviewed the final versions of the material sentences.¹³ Any time intervals deemed implausible were revised until all three judges were satisfied.

3.2.4 Design

Six different presentation lists were used in the questionnaire study. Three of the lists were created so that each event could be evaluated in all three tests without any of the participants evaluating the same event more than once. Then, three additional lists were created in order to counterbalance the positions of *for an hour* and *in an hour* in the forced choice section. Since the three types of test are very different, the trials occur in

blocks. The first block includes 36 *for an hour* versus *in an hour* forced choice questions. The second block includes 36 reading passages each followed by three comprehension questions. The third block includes 36 judgments of sentences beginning with *it took an hour to*. The *it took an hour to* judgment block is last, because it could encourage coerced readings of the events whenever atelic events are presented. The forced choice block occurs first, since this block includes both *for an hour* and *in an hour*. For this reason, it should be less biasing than the other two question types.

The events presented in each block are rotated, as shown in Table 1.

¹³ The three judges were the author, Selene Gardner, and Scott Jackson.

Table 1. Test types in the telicity questionnaire by trial number for the first 3 presentation lists

	Presentation List 1	Presentation List 2	Presentation List 3
Trials 1-36 <i>In/For an hour</i> Forced choice	ate the cupcake, danced, spotted the actress, ...	do the hokey pokey, crossed the finish line, chased the horse, ...	noticed the painting, drink beer, pushed the cart
Trials 37-72 Passages and comprehension questions	notice the painting, drink beer, push the cart, ...	eat the cupcake, dance, spot the actress, ...	do the hokey pokey, cross the finish line, chase the horse, ...
Trial 73-108 <i>It took an hour to</i> naturalness judgments	do the hokey pokey, cross the finish line, chase the horse, ...	notice the painting, drink beer, push the cart, ...	eat the cupcake, dance, spot the actress, ...

With this design, each participant gives one response to each of the events. Across the three presentation lists, each event occurs in each of the three tests. The three presentation lists are repeated as lists 4, 5, and 6, except that the position of *for an hour* and *in an hour* is reversed. Within each of the three blocks, the order of events is randomized in order to lessen the possibility of order effects.

3.2.5 Predictions

The predictions for each verb type in each of the tests are shown in Table 2.

Table 2. Predicted responses for each verb type in each of the three tests

	<i>For/In an hour forced choice</i>	<i>Reading passages</i>	<i>It took an hour to... judgment</i>
Activities (Atelic)	“for an hour”	yes	unnatural (low rating)
Achievements (Telic)	“in an hour” or mixed	mixed	natural (high rating)
Accomplishments (Telic)	“in an hour”	no	natural (high rating)

For the forced choice questions, a higher percentage of “for” choices is expected for atelic events. A higher percentage of “in” choices is expected for telic events. For the achievements, “in an hour” should be predicted since achievements are telic. However, since achievements are very easily coerced, a high number of “for an hour” responses is likely.

For the reading passages and comprehension questions, a higher percentage of “yes” responses is expected for atelic events, especially activities, and a higher percentage of “no” responses is expected for telic events, especially accomplishments. For achievements, the passage will sound a bit unnatural, so the results will probably be mixed.

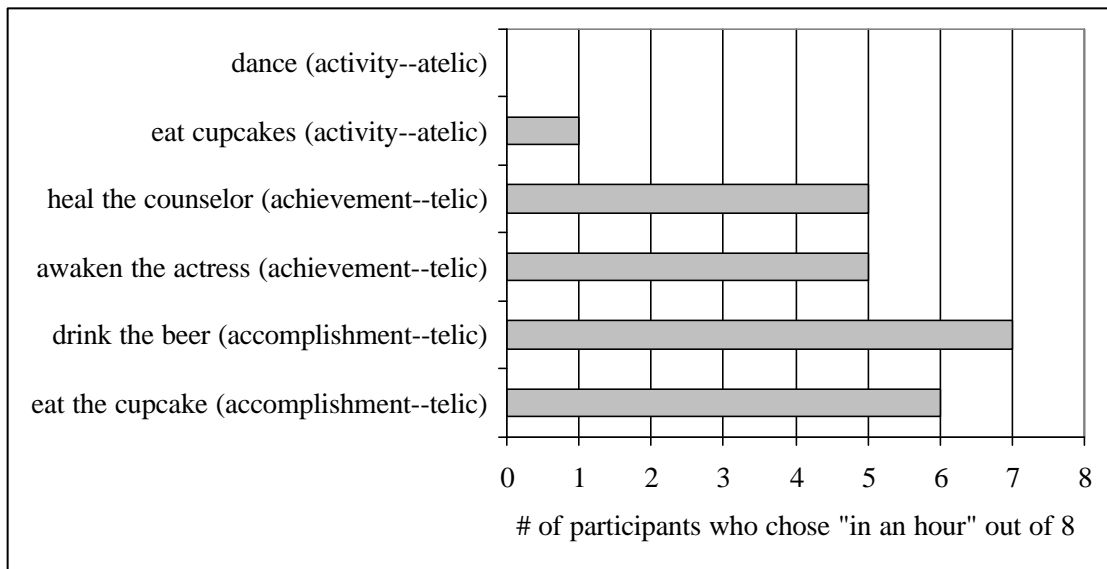
For the “It took an hour to” judgment test, a high rating (very natural) is expected for both types of telic events: achievement and accomplishments. A low rating is expected for activities, because they are atelic.

3.3 Preliminary Results and Discussion

The preliminary analysis that is presented in this chapter is for six of the experimental items. At the time of this report, 24 monolingual English speakers have participated. Since each experimental item appeared in one of three tests for each of the participants, there are eight responses for each item in each test.

The results of the “in an hour” versus “for an hour” forced choice test are given in Figure 2.

Figure 2. Responses from “in an hour” versus “for an hour” forced choice test



All eight of the participants that received each event responded, so the number of *for an hour* responses is equal to the number of *in an hour* responses minus eight. For the event *dance*, there were zero *in an hour* responses and 8 *for an hour* responses. This result

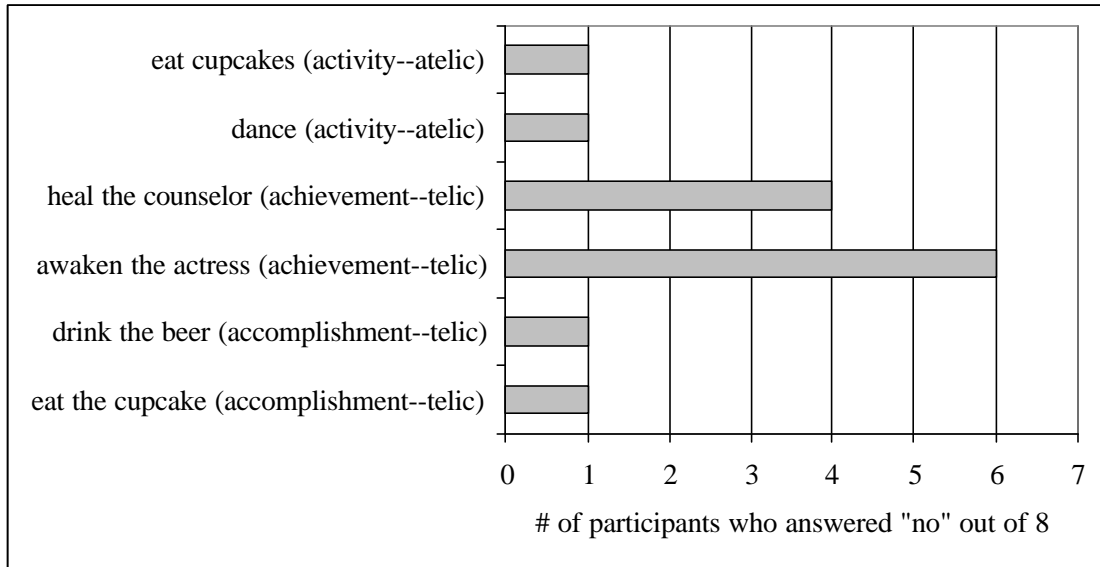
follows the prediction that *for an hour* will be chosen more often than *in an hour* for atelic events.

As predicted, *in an hour* was chosen more often than *for an hour* for telic events, and *for an hour* was chosen more often than *in an hour* for atelic events. For the accomplishments, *drink the beer* and *eat the cupcake*, *for an hour* was chosen only once and twice respectively. This result suggests that the coerced (incomplete interpretation) of these events is available, but as predicted, the adverbial that does not require coercion, *in an hour* is chosen almost all of the time. For the achievements, *heal the counselor* and *awaken the actress*, *for an hour* was only chosen slightly more often. This suggests that the coerced interpretation with *for an hour* is only slightly more available for achievements than for accomplishments. *In an hour* is still preferred the majority of the time, as predicted. *In an hour* was chosen only once for the activity *eat cupcakes*, and it was not chosen at all for the activity *dance*. These results are also as predicted. It is worth noting that there is an acceptable interpretation of both of these events with *in an hour*; it means that the dancing or eating of cupcakes began an hour after a given reference point. Apparently, this interpretation is less available than the non-coerced interpretation that the events have with *for an hour*.

Overall, the results of the *for an hour* versus *in an hour* forced choice test are as predicted. Importantly, in these results, the forced choice test would yield accurate telicity categorizations. With atelic events, *for an hour* was chosen more often, and with telic events, *in an hour* was chosen more often.

The responses from the reading passage comprehension test are given in Figure 3.

Figure 3. Responses to comprehension questions following reading passages



Again, all eight of the participants who were presented with each event gave a response, so the number of “yes” responses is equal to the number of “no” responses minus eight. The questions asked in the comprehension question all had the form, “Did the boy eat the cupcake?” or “Did the girl dance?” A higher number of “no” responses was predicted for telic events than for atelic events.

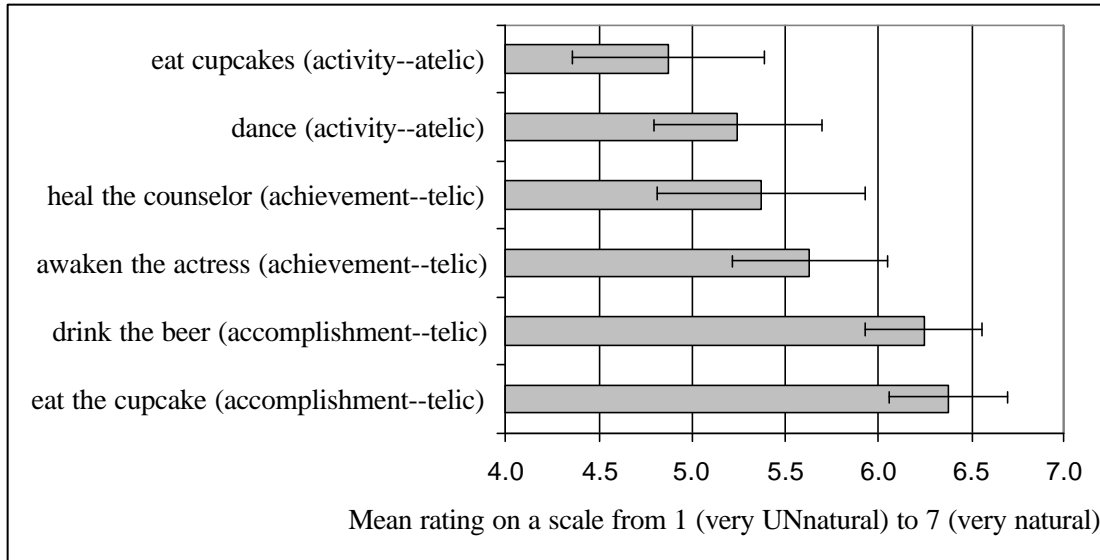
The results were as predicted for the atelic activity verbs, *dance* and *eat cupcakes*. Almost all of the time, the participant indicated that the girl danced even though she stopped in the middle of dancing, and almost all of the time, the participant indicated that the boy ate cupcakes even though he quit in the middle of eating cupcakes. The result for the telic achievement *awaken the actress* was as predicted; there were more “no” responses than “yes” responses. The result for the achievement *heal the counselor* was

unclear, since exactly half of the responses were “yes” and half were “no”. The results for the telic accomplishment verbs, *drink the beer* and *eat the cupcake*, were not as predicted. The responses indicate that in all but one case, the participants did not perceive an endpoint as necessary for the truth of *The man drank the beer* and *The boy ate the cupcake*. This suggests that they interpreted the sentences to mean “The man drank of the beer” and “the boy ate of the cupcake”. These are valid interpretations. In order for this test to be useful for telicity categorizations, such interpretations must be ruled out. One possible way to do this would be to ask *Did the man actually drink the beer?* and *Did the boy actually eat the cupcake?*¹⁴

The results of the *it took an hour* judgment test is given in Figure 4.

¹⁴ Using *actually* or *successfully* in the question was previously recommended to me by Michael Dickey and Andrea Proctor at Northwestern University. I decided not to use these adverbials, because they make the experimental question stand out from the distracter questions. It is now clear that I should take their advice.

Figure 4. Results of “It took an hour to” judgment test



The numbers in Figure 4 are the average of the responses to each event. The bars in Figure 4 represent standard error. All of the numbers between 1 and 7 were chosen on at least one trial except for 1. This means that none of these events in the *it took an hour to* frame were rated as “very unnatural”. A higher rating was predicted for telic events than for atelic events. The results are not as predicted, in that all of the sentences had a mean rating higher than 4, meaning that they are all perceived as natural on average. However, all of the telic events have higher ratings than the two atelic events. The two accomplishments also have higher ratings than the achievements, as predicted. These results suggest that this test might be useful for making telicity categorizations. Analyses of more items is necessary in order to determine whether telic events continue to be rated as more natural than atelic events in this test.

3.4 Conclusions

The results thus far suggest that the *for an hour* versus *in an hour* forced choice test is an adequate diagnostic for telicity. Neither *for an hour* or *in an hour* are chosen 100 percent of the time with any particular event type, but on average, *for an hour* is chosen more often with atelic events, and *in an hour* is chosen more often with telic events.

The results show that the *it took an hour* judgment test successfully differentiates between telic and atelic events. What is lacking is a reliable dividing line between the ratings of the telic and atelic events. Thus far, all of the telic events are rated as more natural than all of the atelic events.

The imperfective paradox reading passage test is apparently an unsuccessful diagnostic in its current form. In a future version of this questionnaire, the adverbs *successfully* or *actually* will be included in the hope that these might rule out an atelic interpretation of accomplishment events in the past tense.

It is important to note that the results of these tests do not answer the question of whether telicity is a categorical or gradient distinction. The availability of coerced and irrelevant interpretations inevitably affects the responses. In order to improve the diagnostic value of any of these tests, it will be necessary to find ways to rule out these coerced and irrelevant interpretations through the details of the test presentation. In the meantime, it appears that the *for an hour* versus *in an hour* forced choice test is the best diagnostic.