What Informs Event Descriptions: Language, Salience, and Discourse in English and Japanese

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of The Ohio State University

By

Hartman Brawley

Graduate Program in Linguistics

The Ohio State University

2012

Master's Examination Committee:

Professor Laura Wagner, Advisor

Professor Shari Speer

Professor Kiwako Ito
Copyright by

Hartman Brawley

2012
Abstract

The aim of this study was to investigate what informs event descriptions. It has been shown that what language one speaks, properties of the event, and properties of discourse all contribute to decisions about what to include in event descriptions. However, little work has been done about the comparative weight of these effects, and how they interact to create trends in the length or the content of descriptions. Previous work has focused on Directed Motion (DM) events, which consist of a moving Figure passing between Grounds. The motion of these events is defined along two dimensions: the Manner, the characteristic movement of the Figure, and the Path, the trajectory along which the movement is carried out. Talmy (1985) proposed a typological dichotomy distinguishing languages which describe Path in the main verb (Verb-Framed languages) and those that describe it outside of the main verb, and Manner in the main verb (Satellite-Framed languages). The typology predicts that information encoded in the main verb will be expressed more, and Whorfian linguistic determinism predicts that these preferences will always be brought to bear. However, subsequent work has suggested that these preferences can be changed if typically dispreferred event components are more highly salient than preferred components. Discourse factors such as addressee identity and visual co-presence may also have an influence. For example, adult speakers should share more knowledge with adult addressees vs. children, and with those that can see the event they are describing vs. those that did not see the event. We presented speakers of English
and Japanese with 14 video clips of Directed Motion events online and asked them to type descriptions of them. The events presented Manner, Path, and Goal elements in either low or high salience, and participants were asked to describe either to an adult or a child, either who viewed the events with them or did not witness the events. The results for description length found effects of discourse, suggesting that discourse factors but not language or salience. The results for inclusion of event components found effects of language, discourse, and salience, suggesting that these influences work cooperatively to inform the content of event descriptions based on the situation.
Acknowledgements

First I would like to thank my project advisor, Dr. Laura Wagner, not just for her guidance and instruction, but also for all of her encouragement along the way. My committee members, Dr. Shari Speer and Dr. Kiwako Ito, also deserve many thanks for their valuable help and suggestions with the project. I would also like to thank Dr. Yuki Hirose and Tokyo University for recruiting our Japanese participants, Sayo Goss for meticulously translating the Japanese data into English, and Dr. JJ Nakayama for his language expertise and assistance with developing the coding scheme. The Departments of Linguistics and Psychology, along with the support from Social and Behavioral Sciences and the Buckeye Language Network, who provided funding for the project, deserve thanks, as well as members of the Developmental Language and Cognition Lab, who assisted with various aspects of the project. Finally, I would like to thank Christine Szostak for listening and helping me talk about my work, my loving parents and sister for all their moral support, and Monte Main for keeping me sane, motivated, and caffeinated.
Vita

2007.................................................................Hilliard Davidson High School

2011...............................................................Poster Presentation, IASCL Conference, Montreal

2012...............................................................B.A. Linguistics, The Ohio State University

Fields of Study

Major Field: Linguistics
# Table of Contents

Abstract..............................................................................................................................................ii  
Acknowledgements............................................................................................................................iv  
Vita.......................................................................................................................................................v  
List of Tables.......................................................................................................................................vii  
List of Figures..................................................................................................................................viii  
Introduction.........................................................................................................................................1  
Methods.............................................................................................................................................22  
Results...............................................................................................................................................32  
Discussion..........................................................................................................................................44  
References...........................................................................................................................................53
List of Tables

Table 1. Stimuli Salience Configuration Movie Types .......................................................... 26
Table 2. Discourse Conditions .............................................................................................. 39
Table 3. Example Coded English and Japanese Descriptions .............................................. 33
Table 4. Results of the Subjects Analysis ANOVA .............................................................. 36
Table 5. Results of the Items Analysis ANOVA .................................................................. 36
List of Figures

Figure 1. Component Inclusion Scores Across Languages.................................38
Figure 2. Number of Sentences Across Discourse Conditions...........................40
Figure 3. Component Inclusion Scores Across Addressee Conditions....................41
Figure 4. Component Inclusion Scores Across Co-Presence Conditions...................42
Figure 5. Component Inclusion Scores Across Salience Configurations..................45
Introduction

One of the most common objectives of language is to relay information about events. Events may either be observed, understood from hearsay, or imagined. Especially in the first case, it is reasonable to expect a correspondence between entities in the world and the descriptions that reflect them. If the description is to be understood by the hearer, it should convey actual content about what occurred. However, the correspondence between event and description cannot be one to one, for several reasons.

First of all, since there exist different systems of encoding events (i.e. distinct languages), descriptions across these systems will naturally be different. Cross-linguistic differences are not completely arbitrary; they reflect trends about what content is canonically or most frequently described by speakers. For example, Talmy (1985) proposes a typological dichotomy that groups languages based on their expression of motion in events. Different languages more easily encode different event components according to their type within the dichotomy. Whorf's linguistic determinism hypothesis takes this one step further by claiming that these patterns will predict what is encoded by speakers in event descriptions. Secondly, descriptions are tailored to the situations in which they are uttered. Any particular description takes many discourse factors into account, such as relative social standing and shared information with an addressee. A description spoken to a young child for example may show stylistic differences, include repetitions, and be simplified on the part of the speaker in an attempt to facilitate
understanding (Clark and Bernicot, 2008; Fernald and Morikawa, 1993). Furthermore, information that is thought to be understood by the addressee, especially after it is mentioned once, may be simplified or even omitted from further mention (Brennan and Clark, 1996). Finally, there is no true description of an event. To help explain this, consider that events themselves do not exist as discrete entities. They have no intrinsic beginning or end, and they consist of infinitely many factors, none of which freely lend themselves to be either included or omitted in a description. This means that there are infinitely many possible descriptions of an event, all of which accurately reflect it to some degree. The properties that are ultimately described are subject to the human perception of the observer. That said, salience must also play a role in the decisions made by a speaker on what to describe. As events are perceived and formed into event representations, pieces of events can be more or less prominent or imageable than others. More interesting components are more likely to be attended to, and therefore more likely to be encoded in a description (Hickmann et al, 2009).

All of these factors influence description, and the extent to which they are independent or interact with one another is the subject of this thesis. These factors have each been investigated individually in research designs, but little work has been done using a design that manipulates all of these factors at once to examine their relative weight in informing descriptions. We aim to accomplish this by analyzing descriptions of events in two different languages. The following sections refer to three broad categories of factors influencing description: language, discourse, and salience. These
categories are very broad, and therefore will not be covered exhaustively in this paper. Instead, we will select and examine a few specific influences from each category, and present them in each of the following sections. First we discuss the role of language: differences between languages’ expressions of space and motion. Then we discuss the role of discourse: information status and some basic discourse conditions. Then we discuss the role of salience: relative salience of event characteristics and how cognition facilitates event apprehension and description formation. Finally, we will outline the structure and predictions of the current study. From there, we will present our Methods, Results, and Discussion.

Language Factors

In order for an event to be described in a specific language, a speaker needs to analyze it cognitively and create an event representation. These two processes need not happen one after the other, however. Gleitman et al. (2007) demonstrated in an eye-tracking study that gist apprehension and linguistic planning can both occur within the first 200 milliseconds (ms) of viewing an event, suggesting that non-linguistic and linguistic analyses happen cooperatively and in unison. They presented participants with a black box for only 60-75 ms on either the left or right side of the screen followed by a scene involving two subjects, one on either side, that they described. The box directed their gaze unconsciously to one of the two figures at onset of the scene, therefore making it more salient at first. The scenes consisted of an action carried out by an agent onto a
patient (for example, a dog chasing a man). The prediction was that if language and scene apprehension happen separately, then participants should describe the scenes according to the highest frequency linguistic patterns, such as agents before patients, active voice over passive voice, animates mentioned before inanimate objects, etc (in the example case, "a dog chased a man" is higher frequency than "a man was chased by a dog"). Even though participants reported not noticing the black box, it successfully guided them to describe the scene beginning with the figure in the location that the black box had appeared in just before the trial. Furthermore, participants described the figure to which their attention was directed even when it resulted in the lower frequency description ("a man was chased by a dog").

These results suggest that participants were gaining an initial "gist" of the scene at the same time they were formulating their description, both within 200 milliseconds of viewing the event. They also suggest that the non-linguistic and linguistic analyses may have an influence on one another. If they do, then speakers of different languages may form essentially different event representations, and whether this is true is the subject of much debate. Language must always appeal to cognition, as a description is formed from an event representation. Furthermore, even in different language communities we expect speakers to share perceptions about events, to the extent that they share similar sensory organs. However, to what extent does the structure of one’s language shape the formation of event representations, to facilitate ease of linguistic encoding? The Sapir-Whorf linguistic determinism hypothesis offers a strong claim about the relationship
between language and thought: native speakers of a language are compelled to attend only to those dimensions of their experience that suit the syntactic categories of their language and ignore those that do not (Whorf, 1956). For example, Lucy (1992) demonstrated that speakers rely on their particular language's existing semantic categories to categorize novel objects. Adult speakers of English and Yucatec Maya were presented with 17 triads of objects, one original and two alternates: one that different in its shape and one that differed in its material composition. An example triad would be ceramic bowl (original), a ceramic plate (shape alternate), and a metal bowl (material alternate). Participants were asked to decide which alternate was more similar to the original object. English speakers reliably chose the material alternate, and the Yucatec Maya speakers chose the shape alternate. Lucy concluded that because English forms object contrasts on the basis of shape more than material composition, and Yucatec Maya material more than shape, that these linguistic habits drove participants to find distinct salient differences among the same objects. This view of language affecting thought in any situation regardless of context we will refer to as the Global perspective.

There is another possible explanation of the relationship between language and thought, described by Gennari et al. (2002) as the Language as Strategy Hypothesis. It claims that speakers can rely on the way they prepare to use language to help them perform certain non-linguistic thought tasks that rely on similar demands as language. Gennari et al. demonstrated this by presenting English and Spanish adults with 36 videotapes of DM events. Participants were in one of three conditions, one where they
had to describe each event, one where they did not, and one where they instead were asked to repeat nonsense syllables after each event. Then, all participants were shown 108 DM events, 36 of which they were shown previously, and were asked to indicate for each event whether they had seen the event before. Lastly, they provided similarity judgments on the same set of events. They found that speakers of either language that described beforehand did not recognize events better according to the information they encoded in their descriptions. However, they did judge events with the same information encoded in their descriptions as similar more often. Therefore, only in executing certain tasks can the speaker benefit from relying on their linguistic perspective. Furthermore, a direct appeal to the linguistic perspective can change non-linguistic cognition.

Papafragou and Selimis (2010) asked Greek and English speaking participants to view 48 motion events, describe them, and then match each one to one of two other events, each of which varied by either the Manner or the Path. Participants categorized them according to verb lexicalization patterns in their native language - English speakers grouped events with similar Manners and Greek speakers with similar Paths. However in a second experiment participants were not asked to describe the events and all linguistic cueing was removed from the procedure. In that case the effect did not appear. This finding demonstrates the effect of the language in the instructions and descriptions in the first task contributing to performance in a subsequent, non-linguistic task. This strategy formation based on the presence of linguistic cueing we will refer to as the Local perspective.
Both Global and Local effects may influence event apprehension and
descriptions. If both kinds of effects are present, we would expect to see a baseline effect
from language onto thought when no direct appeal to language is made, and a
strengthening of the effect when one is. However, the current investigation does not use
a non-linguistic measure, and the only thing being examined is descriptions. Therefore,
the only kind of thought required of participants is the planning required to form an
utterance, what Slobin (1996) refers to as “thinking for speaking.” Another way of
thinking about the Global perspective is that descriptive tendencies of a language should
hold even when other non-linguistic factors such as discourse situations and salience
differ. If Whorf was correct and language affects thought all the time, then language-
specific rhetorical styles should predict what is included in descriptions even when other
factors such as discourse situations and salience of event components vary. We will now
discuss how languages systematically differ in their characterization of space.

**Directed Motion Events and the Talmy Typology**

As previously mentioned, Talmy (1985) proposed a typological dichotomy for
languages’ representations of space, specifically Direct Motion events (henceforth DM
events). DM events are defined as a Figure, a body in motion, travelling between
Grounds, referential points in space. The motion undertaken by the Figure is
characterized by two dimensions: Path of motion and Manner of motion. The Path of
motion is the trajectory that the Figure follows during the motion event. Manner refers to
the characteristic, intentional, or idiosyncratic features of the motion. For a simple example, consider the difference between walking and jumping. Both of these unique motions could be undertaken at a fixed location in space, without a trajectory. When the Manner travels along a Path, Directed Motion takes place. Grounds, the physical locations between which the Figure travels, often coincide with physical objects, but they do not have to. There are three basic categories of Grounds. The first two are Sources and Goals. A Source is where the Figure begins its motion, and a Goal is where the motion ends. Note that the Path trajectory of the motion connects the Source and Goal of the event. Grounds belonging to the third category exist along the Path of the motion, and past which the Figure travels, but the motion neither begins nor stops at these points. Talmy’s typology organized languages into two groups based on their syntactic means of Path relative to Manner, and what is prototypically encoded by speakers. These two types are:

1. Verb-Framed languages: which encode Path in the main verb and Manner outside.
2. Satellite-Framed languages: which encode Manner in the main verb and Path outside.

Satellites are defined as peripheral structures outside of the main verb. Talmy provided examples of these as particles or prepositions (common in English), verb prefixes (as in Russian and Latin), and non-head elements of verb sequences (as in Tibeto-Burman languages). The typology claims that Verb-Framed languages will have
more Path verbs and will encode Path information more frequently and with greater ease than Satellite-Framed languages. Satellite-Framed languages typically encode Manner in the main verb instead, and are therefore predicted to have more Manner verbs and encode Manner more frequently and with greater ease. Support for this typology has been found in many different languages, notably through the body of work done on studying "frog stories," a term first used by Berman and Slobin (1994) referring to elicited descriptions of illustrations in the 1969 Mercer Mayer children's book, *Frog, Where Are You?* This book was considered appropriate for eliciting descriptions of DM events because the book contained no words and depicted dynamic motion events. Subsequent work found expected typological trends within elicited descriptions in many different languages, such as Spanish, French, Turkish, Italian, Hebrew, Dutch, German, English, Mandarin, Thai, Russian, and Japanese, and others. The present study investigates English, a Satellite-Framed language, and Japanese, a Verb-Framed language. We turn now to the typological and other syntactic contrasts between these languages.

*English and Japanese*

Consider the following example description in English (1) and Japanese (2).

(1) The boy ran from the house to the street.

In this English example, the Figure of the DM event is *the boy*. Two Grounds are
mentioned, a Source and a Goal. The Source is encoded by *from the house*, where the boy began moving, and the Goal is *to the street*, where the boy finished moving. The Manner of the motion is encoded by the action verb *ran*, which is the main verb. The Path is implied by mention of the Source and Goal, but no trajectory is explicitly mentioned. Now look at the Japanese equivalent:

(2) Otokonoko-wa uchi kara michi-ni hashitte itta.

(boy-GEN house from street-LOC running went)

“The boy ran from the house to the street.”

In this sentence, the same elements are represented. The Figure, *otokonoko*, "the boy," the Source, *uchi kara*, "from the house," and the Goal, *michi made*, "to the street," are all expressed similarly to the English example, but the Manner and the Path are expressed differently. Rather than in the main verb, the Manner is expressed by the gerund form *hashitte*, "running." The Path is expressed by the main verb *itta*, "went." Note that by contrast *went*, the English equivalent of *itta*, does not express Path information. The Japanese verb *iku*, “to go,” expresses deictic information about the relative trajectory of the subject of the verb; namely, that the subject was located at a point in space and it moved away from that point. This “away” component to the meaning is especially apparent when *iku* appears without Goal specification. Consider the difference in meaning between (2) and (3):
(3) Otokonoko-wa hashitte itta.
   (boy-GEN running went)
   “The boy ran away.”

In the case of (3), *itta* no longer means “went” but is more closely translated as “away,” expressing deictic Path information about the subject “the boy.” Tsujimura (2002) claims that in order to properly frame a DM event in Japanese, Path must be expressed somewhere in the clause. To illustrate this, consider the following examples:

(4) *Otokonoko-wa kooen-ni        aruita.*
    (boy-GEN park-LOC walked)
    “The boy walked to the park.”

(5) Otokonoko-wa kooen-made aruita.
    (boy-GEN park-until walked)
    “The boy walked to the park.”

(6) Otokonoko-wa aruite kooen-ni        itta.
    (boy-GEN walking park-LOC went)
    “The boy went to the park walking.”

(4) is not licensed because there is no Path information conveyed. *Aruita* is a Manner verb, and -ni is a Goal marker, it does not encode information about the trajectory of motion. Thus, the meaning of (4) is actually closer to “He walked in the park.” By replacing -ni with -made, a Path and Goal marker similar to English “until,” (5) successfully describes directed motion. Finally, (6) demonstrates that -ni can be used felicitously in a DM event description, but only when used with a Path verb, such as *itta.* Also notable in (6) is that, unlike (4), either -ni or -made are acceptable here.
This case supports both the typological claim that Path-Framed languages encode more Path information in verbs and that they encode more Path information overall. Path verbs in Japanese are high frequency, especially as the heads of clauses. These include *agaru* and *noboru*, “to ascend,” *kudaru* and *sagaru*, “to descend,” *iku*, “to go,” *kuru*, “to come,” *kaeru*, “to return,” *idoosuru*, “to move in space, to transfer” and *susumu*, “to proceed” (Beavers, 2008; Beavers et al., 2009; Tsujimura, 2002). By contrast, English Path verbs such as *ascend*, *descend*, *traverse*, *transfer*, and *proceed* are infrequently chosen to describe an event, in favor of Manner verbs such as *walk*, *jump*, *careen*, *roll*, etc., accompanied by a directional particle such as *up*, *down*, *across*, etc., expressing the Path. The case of (3) also demonstrates that forms in different languages that seemingly express the same meanings may entail different aspects of an event. This is an especially important consideration in studying cross-linguistic encoding of events given that any investigator comes from the inherently biased perspective of their native language. The investigator must therefore be sensitive to the intuitions of the native speakers of the languages s/he is studying.

It is important to note that the typological patterns described in 1. and 2. are not obligatory to create grammatical descriptions of events for speakers of either language type, they represent prototypical patterns (Hickmann et al, 2009). Natural descriptions of events can be uttered that do not conform to typological expectations but these are generally less frequent, and sometimes require greater syntactic marking, and are longer than their typologically prototypical counterparts. Consider (4) and (5), descriptions in
English:

(4) Tre walked up the stairs.
(5) Tre ascended the stairs.

(4) is a typologically prototypical utterance. Being a Satellite-Framed language, English is expected to express Path information outside of the main verb, and in (4) it is expressed by the preposition _up_. In (5), the verb _ascended_ defies the Manner verb expectation by describing the trajectory of Tre in the main verb. First Tre was at the bottom of the stairs and then he was at the top. Now let us look at Japanese:

(6) Mayuki-wa tobi-notta.
    (Mayuki-GEN jump-ascended)
    “Mayuki jumped up.”
(7) Mayuki-wa tonda.
    (Mayuki-GEN jumped)
    “Mayuki jumped.”

(6) shows a typologically predicted Japanese descriptions. _Tobi-notta_, the clause-final verb compound, is made of a bound morpheme, _tobi_, the verb root from “to jump,” and a free morpheme _notta_, “ascended.” The final morpheme of this common compound structure also carries the inflection, and importantly is usually a Path verb. Using only a Manner verb is also possible in Japanese as in (7), but they do not describe Path information. In order for the Path to be expressed, the Path verb must come last, either as the final morpheme in a verb compound or as its own verb. Furthermore, the Manner-
Path ordering of the verb compound morphemes cannot be reversed, as shown in (8):

(8)  *Mayuki-wa nori-tonda.
    (Mayuki-GEN ascend-jumped)  
    “Mayuki jumped up.”

(9)  Mayuki-wa notte tonda.
    (Mayuki-GEN ascending jumping)  
    “Mayuki ascended and then jumped.”

(8) is not grammatical because the Path morpheme in the compound comes first, and then the Manner compound. If the gerund form is used, as in (9), the sentence is grammatical but no longer describes a single, integrated action as (6) does with tobi-notta, “jumped up.” (8) and (9) suggest that in Japanese, whose clauses are head-final, Path encoding is also placed at the end of the clause. If further verb morphemes describing Manner follow the Path verb, it suggests that a new event is being described. Thus, the description in (9) describes two, sequential events. These facts support Talmy’s assertion that in a Verb-Framed language Path is more prominent than Manner and is more integral to event descriptions.

Comparing descriptions of English and Japanese, based on their typological differences we predict that overall component inclusion will differ. Based on the typology, we expect Japanese to encode more Path than English, and English to encode more Manner than Japanese. We do not expect Goal to be encoded differently by speakers of different languages.
Discourse Factors

As Slobin (2004) points out, typology alone cannot be the sole account for what guides languages use because it does not take discourse factors into account. The typology outlines a rhetorical style, or a basis for organizing dimensions of the event for description, but the role of paralinguistic knowledge is crucial. A discourse refers to a linguistic exchange, accompanied by the many extra-linguistic discourse factors that accompany and guide it. These factors include shared knowledge (both inside and outside of the immediate conversation) among speakers, information about addressees, and context. Collectively, discourse factors are like a subconscious log kept among individuals containing all of the relevant facts and inferences necessary to generate appropriate linguistic contributions. The log is incrementally updated as the discourse develops. Kamp (1981) refers to this log as a Discourse Representation Model (DRM). DRMs are not truly shared among speakers, each speaker maintains their own DRM which can be inferred, but not read, by other speakers. Contained within a subset of the DRM is all of the shared expectations and knowledge speakers with addressees, known as common ground (Clark and Schaefer, 1989; Schober and Clark, 1989). Common ground contains both inferences made about speakers prior to the linguistic exchange (based on perceptions about their physical and social qualities, as well as prior experiences with speakers if any), and the implicatures of utterances during the actual exchange. Along with these inferences, a speaker takes into account general world knowledge, shared culture, and past experiences if any that are shared with her/his addressees. This process
occurs not just before, but incrementally throughout an exchange as input continues to be received from subsequent utterances (Brown-Schmit and Hanna, 2011).

Speakers generally engage in a discourse with the objective of making it a success; that is, to be understood and convey information to one another. Clark and Schaefer (1989) outlined three general means by which speakers achieve this objective. First, speakers assume a certain amount of common ground at the onset of the exchange. Second, throughout the discourse, they attempt to add to their common ground. Finally, speakers add to common ground by uttering appropriate things at appropriate times. Clark and Schaefer argued that in order to fulfill this third goal speakers weigh many extra-linguistic discourse factors against one another to inform the language planning process. This is reflected in findings demonstrating that speakers tailor their utterances to suit the needs of addressees in an exchange, presumably to facilitate understanding by their addressees. Two factors that have shown this effect are addressee identity and visual co-presence.

Many different qualities about interlocutors may lead to accommodations in descriptions, but perhaps one of the most studied qualities is relative age, speaking to adults vs. children. Not only do adults generally assume that they have more common ground with adults than with young children, but in most situations they believe children know less about the world than they do. This may lead to simpler and shorter utterances, along with vocal changes such as pace, pitch height, and pitch contour. Fernald et al.
(1989) studied motherese or infant-directed speech in men and women of six languages (American English, British English, French, Italian, German, Japanese). Similar features of child directed speech were observed in all languages, in including slower-paced talking, greater pitch height, greater pitch movement, and longer pauses, suggesting that all languages attended to perceived needs of the infants they were addressing. The features of child-directed speech may be culture or language-dependent, however. Fernald and Morikawa (1993) found in their study of Japanese and English speaking mothers talking to infants of 6, 9, and 12 months that speakers of different languages showed different patterns. 30 mother-infant pairs of each language were observed and recorded in their homes. American mothers spoke less overall to their infants, but used more object noun labels than their Japanese counterparts. Fernald and Morikawa (1993) attributed this to a cultural difference causing the two mother groups to have different beliefs about the addressee needs of their infants. Japanese mothers had less of an expectation for the baby to be independent, and used more terms promoting social cohesion, which is prevalent in Japanese culture. American mothers on the other hand valued education, and used less phrases of social cohesion. These findings suggest that, while there might be language-specific and cultural-specific patterns, speakers believe that adults have different needs from an utterance than infants, and they tailor their utterances accordingly. While the studies reviewed here studied infant-directed and not child-directed speech, we expect the same trend of adult speakers speaking less to younger-than-adult speakers.
Returning to visual co-presence, the term refers to speakers sharing visual access to an event. Under these conditions, a speaker can assume that s/he shares more knowledge about the event with her/his audience. Therefore, more salient information may be dropped from mention (Clark and Schaefer 1989). In their Tangram assembly study, Schoeber and Clark (1989) demonstrated that visual co-presence and participation in the discourse has a profound effect on the utterances in a discourse and the apprehension of events. 14 pairs of students assembled Tangram figures while 14 overhearers who were not visually co-present attempted to assemble the same figure based on their instructions. Both when listening to a tape of the conversation (Experiment 1), and actually being present (Experiment 2), overhearers were much slower and poorer at constructing the Tangrams. Schoeber and Clark theorized that this was because the two cooperating participants were visually co-present and could refer to the figures they were constructing.

Visual co-presence also affects early choices about description planning. Lockridge and Brennen (2002) suggested that speakers do not always tailor their utterances for addressee's needs, but especially do so when they believe their addressees share less information about the event via visual co-presence. They modeled Brown and Dell (1987)'s study in which participants were shown pictures of events, each accompanied by a short written description. Participants were asked to read the events and then describe them to confederate addressees, who had not see the cards. The events contained either a typical (e.g. a knife) or an atypical (e.g. icepick) instrument for the
action (e.g. stabbing), and found that participants described atypical instruments more overall. Lockridge and Brennan were concerned that Brown and Dell did not take visual co-presence into account; participants all described events for someone who had not seen the event. Lockridge and Brennan modeled their study by placing 72 pairs of participants in one of three co-presence conditions: no co-presence, similar to Brown and Dell's experiment in which only one participant could see the event, partial co-presence, where both participants could see the event on separate screens but not each other, and full co-presence, in which both participants could see the event on a single screen and each other. The study found that participants described atypical instruments more in the not co-present condition (replicating Brown and Dell's findings), but at roughly equal levels in either of the co-present conditions. Visual co-presence makes a difference to speakers when planning their descriptions. Overall, we expect that the discourse factors associated with addressee identity and visual co-presence will affect lengths of descriptions, but not component inclusion for speakers of either English or Japanese.

Salience Factors

In addition to factors of speakers and addressees, properties of events themselves guide descriptions. When a speaker views an event, certain components stick out and are more likely to be integrated into a speaker’s event representation. Therefore, in general, event components that are highly salient should be described more than event components that are less salient (Hickmann et al, 2007). However, what makes event
components perceptually salient? In the interesting case of Manner, Slobin (2006) pointed out that Manner is actually a category encompassing many different dimensions of motion, “including motor pattern (e.g., hop, jump, skip), often combined with rate of motion (e.g., walk, run, sprint) or force dynamics (e.g., step, tread, tramp) or attitude (e.g., amble, saunter, stroll), and sometimes encoding instrument (e.g., sled, ski, skateboard), and so forth.” How complex these dimensions are contribute to the inherent salience of Manner within the event. However, there is another dimension of salience: how readily encodable the information is for a language. According to the Talmy typology, Satellite-Framed languages, whose speakers encode Manner in the main verb, find Manner more readily encodable than Path, and the opposite is true for Verb-Framed languages, whose speakers encode Path in the main verb. But how does this interact with the inherent salience of components within the event? If Manner is less salient, will Satellite-Framed language speakers be less likely to omit it from descriptions because it is encoded in the main verb? Will Verb-Framed language speakers omit it more readily?

Recall that the Path of a Figure is its trajectory of motion. The more complex the trajectory of an event is, the more salient it is to the viewer. For example, if a Path is inferable from the Source and Goal of the event (a straight line), it may be omitted from mention, especially in a Satellite-Framed language. Consider the English description in (1): The boy ran from the house to the street. While the Source and Goal are both specified, the Path is only implied. If the phrase in a straight line were suffixed to this description, it would specify the Path. If the Path is complex and not inferable from
Grounds, it may license more elaboration from the describer. Certain spatial dimensions upon which the Figure travels may be more salient than others. For example, a Figure with a Path containing a vertical component is more interesting to describe than a Figure traveling across a flat plain.

Goal salience is somewhat less nuanced than Manner or Path. What licenses mention of an event’s Goal, where the Figure ended its motion, is the presence of a nameable object whose location coincides with the termination of the motion. For example, consider two events, a ball rolling to the center of a table, and a ball rolling to a cup in the center of the table. The cup is a more salient Goal than the empty center spot of the table, and is more likely to be mentioned than the center of the table. A Figure having an intentional (e.g. the ball rolled into the goal) or instrumental (the key slid into the lock) relationship with a Goal leads to higher Goal salience as well.

Current Study

To investigate the various possible influences on event descriptions, we presented participants of English and Japanese with videos of DM events online and asked them to type descriptions of each event. Participants were placed in one of four discourse conditions, in which they were asked to describe the events either to an adult or a child, who either was watching the videos with them or was seated across the room and could not see the movies. Furthermore, we manipulated the movies themselves to present different salience configurations of Manner, Path, and Goal to participants. All
components were either in high or low salience. From each of participants’ descriptions we recorded whether they encoded Manner, Path, or Goal information, and how many sentences they used. With this design we were able to examine effects of language, discourse, and salience, and their interactions.

Based on prior work and reasoning, we expected certain findings. Based on the DM typology, we expected that Japanese speakers would describe more Path than English speakers, and the English speakers would describe more Manner than Japanese speakers, but that Goal inclusion would be the same for both languages. We expected the descriptions of Japanese and English overall to be of about the same length. For discourse factors, we expected participants to use more sentences when describing to adults, and to not present addressees. We did not expect different inclusion of Manner Path and Goal based on the discourse situation. For the salience configurations in different movies, we expected for both languages that highly salient components would be described more than low salience components, and that this would interact with the expected language finding, based on typological patterns of component inclusion. The nature of this interaction between language and salience is potentially informative for the Whorfian hypothesis. Specifically, we expect that if the Whorfian hypothesis holds and speakers attend to certain dimensions of their experience regardless of the context, then English speakers should describe Manner and Japanese speakers should describe Path regardless of the salience of the components in the event. Another way of saying this is that if Manner is more prominent in English, then English speakers should describe
Manner both for movies in which Manner is presented in high or low salience. In turn if Path is more prominent in Japanese, then Japanese speakers should describe Path whether it is in high or low salience. Speakers of languages for whom a component is not as syntactically prominent (e.g. Goal for both language types) should describe that component more often when it is presented as highly salient than when it is in low salience. If the Whorfian hypothesis does not hold, speakers of both languages should show this salience-sensitive inclusion pattern for all components. We did not expect any other interactions.

Methods

Participants

Seventy English speaking undergraduate students from Ohio State University in Columbus, Ohio and 71 Japanese speaking undergraduate students from Tokyo University in Tokyo were recruited to participate in the study. In the English group, 38 students were female and 32 were male, and in the Japanese group 18 students were female, 52 male. All students reported to be monolingual, apart from language classes they had taken in school. English students were part of a Psychology research pool and received half of a Research Experience Program credit to fulfill a course requirement for their participation. Japanese students received 1000 Yen. English-speaking students were recruited using a departmental research posting website in Psychology, and the
Japanese-speaking students were recruited using flyers in classes taught at Tokyo University by our collaborator, Yuki Hirose, who helped us recruit participants abroad.

**Materials**

When all students were recruited they were directed to an online survey constructed on a series of web pages on Surveymonkey.com. Each page contained a link to one of 14 short movie clips of DM events. The clips were each about 3-6 seconds in duration, and were cut using Final Cut Pro software. The audio track was removed from all movies; none of them contained sound. All videos involved a person's hand moving a stuffed animal on one of three different kinds of environments, each of which was a foam core ramp on a table. One board was a flat strip, the second was a larger square with a Lego tree in the back left corner and a Lego house in the back right corner, and the third was a two-platform environment with a ramp connecting the two levels. In some cases this board also contained two plastic baby food bowls, one on each level. The bowl on the top was red, turned open side-down, and to the right of center, and the bowl on the bottom was orange, turned open side-up, and to the left of center. Videos were created to place three DM event components, Manner, Path, and Goal, in different configurations of salience. Table 1 presents all 7 of the salience configurations used in the movies, along with images of a sample event from each configuration. The images are the last frame of the directed motion, to illustrate the Goal of the sample event, along with annotations to illustrate the Manner and the Path of the motion underwent by the Figure.
In each configuration, one, two, or all three of the event components are highlighted, following the prediction that components that are highly salient will be described more often than non-highlighted components. It is also important to keep in mind that non-highlighted components are still somewhat salient, just less than highlighted ones. Any DM event contains a Path, Manner, and Goal, but they may not be all equally salient in an event. In the case of Manner for our events, in salience configurations where it is highlighted the Figure undergoes motion that is idiosyncratic and intentional, namely either hopping or sliding. Highlighted Manners are expected to be more highly salient, and therefore encourage participants to encode Manner in their descriptions. In configurations where Manner is not highlighted, the Figure simply moves in a very direct trajectory toward the goal without intentionality, and this movement is expected to produce descriptions containing non-specific or no expression of Manner. For Paths, in highlighted configurations the Figure moves either up or down the ramped board, where the vertical trajectory is expected to be more salient Path than in the non-highlighted cases, in which the Figure moves across (either to the left or to the right) a flat plane. Since Paths are trajectories leading to Goals, the relative salience of Goals may also interact with Path salience. That is, Path is expected to be described more when it is highlighted, but it may even be described more in an event where it is not highlighted, but the Goal is. For Goals, in highlighted configurations a Ground exists on the board where the Figure ceases its movement. On the flat board used for configurations where Path is not highlighted, these Grounds are a Lego tree and a Lego
house. On the board with the ramp, used for configurations where Path is highlighted, these Grounds are an upside-down and a right-side up baby food bowl on the lower and upper platform connected by the ramp. When Goal is not highlighted, the Figure either moves across the board, or up or down the board if Path is highlighted.

<table>
<thead>
<tr>
<th>Salience Configuration</th>
<th>Salience of Event Components</th>
<th>Event Synopsis</th>
<th>Sample Event Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>Manner – high salience Path – low salience Goal – low salience</td>
<td>The animal turns to the left, hops to the left side, and faces forward.</td>
<td><img src="image1" alt="Sample Event Image 1" /></td>
</tr>
<tr>
<td>P</td>
<td>Manner – low salience Path – high salience Goal – low salience</td>
<td>The animal moves from the bottom of the ramp to the top of the ramp.</td>
<td><img src="image2" alt="Sample Event Image 2" /></td>
</tr>
<tr>
<td>G</td>
<td>Manner – salience Path – salience Goal – high salience</td>
<td>The animal moves from the front center of the board to in front of the tree in the back left.</td>
<td><img src="image3" alt="Sample Event Image 3" /></td>
</tr>
<tr>
<td>Configuration</td>
<td>Manner – high salience</td>
<td>Path – high salience</td>
<td>Goal – low salience</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------</td>
<td>----------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td><strong>MP</strong></td>
<td><strong>High</strong></td>
<td><strong>High</strong></td>
<td><strong>Low</strong></td>
</tr>
<tr>
<td><strong>PG</strong></td>
<td><strong>Low</strong></td>
<td><strong>High</strong></td>
<td><strong>High</strong></td>
</tr>
<tr>
<td><strong>MG</strong></td>
<td><strong>High</strong></td>
<td><strong>Low</strong></td>
<td><strong>High</strong></td>
</tr>
<tr>
<td><strong>MPG</strong></td>
<td><strong>High</strong></td>
<td><strong>High</strong></td>
<td><strong>High</strong></td>
</tr>
</tbody>
</table>

Table 1: All event types used in the stimuli video clips of DM events. The salience of event components is provided for each configuration, along with a synopsis and an image from a sample event belonging to each configuration.
Procedure

All participants completed the experiment online on a survey designed using Surveymonkey.com. They were first instructed to imagine one of four discourse scenarios, corresponding to those displayed in Table 2, and keep it in mind while they described all movies. The conditions were based on two discourse factors: visual co-presence and addressee identity. In two of the conditions, participants were asked to imagine that they were describing the movie for someone who was watching the movies with them and could see everything that was happening. In the other two, participants were to describe for someone sitting across the room who could not see the events. Also, two of the conditions asked participants to imagine describing the movies for an adult, and the other two asked them to describe to a child.
29

Visual Co-presence | No Visual Co-Presence

| Addressing an Adult | Describe the movies for an adult who is sitting next to you and watching the movies with you. | Describe the movies for an adult who is sitting across the room and cannot see the movies. |
| Addressing a Child | Describe the movies for a child who is sitting next to you and watching the movies with you. | Describe the movies for a child who is sitting across the room and cannot see the movies. |

Table 2: Instructions based on discourse conditions presented to participants at the onset of the experiment.

After receiving instructions containing one of these discourse scenarios, participants viewed 14 pages, one at the time, each one containing a link to a DM event and a box in which they were to type their description of the event. At this stage, no further instructions guiding the content of their descriptions were given. Participants were told that they could watch each movie clip as many times as they wanted before describing the event. Every participant saw two events from each of the seven salience configurations, making 14 events in all. The events were randomly ordered into one list, and a second list was created out of the reversed order of the first list. This created 8 conditions in all: four discourse scenarios across two lists. The Japanese participants
were run first, and the English participants were balanced according to the number of Japanese participants in each condition. 36 English speakers were presented with the forward order, and the remaining 34 English speakers saw the events in the reverse order. For Japanese speakers, 36 saw the events in the forward order, and 35 in the reverse order. Each participant provided 14 descriptions, with 9 exceptions. Each of these 9 participants (4 Japanese, 5 English) skipped exactly one description, therefore providing 13 each. After watching and describing all 14 events, speakers were asked to complete a short questionnaire. They were asked to report their first language, their sex, what other languages they had experience with and to what extent, how many children they had, and finally the amount of experience they had talking to children. After the Japanese data was collected it was glossed and translated into English by a hired native speaker, and then coded primarily on the basis of the glosses, which preserved the grammatical structure of the descriptions.

Coding

Linguistic Quantity

To gauge whether discourse conditions or salience configurations influenced how elaborative participants’ descriptions were, average number of sentences per description and average number of event components per description were recorded for each subject. Sentences were counted by the number of periods separating text, as included by the participant within the description, within reason. This means that, as some participants
used a more conversational style with their descriptions, run-on sentences without periods were not prescriptively divided into multiple sentences. We attempted to preserve the intended flow of the description as expressed by the participant. On the other hand, single sentence descriptions containing no period were still counted as one sentence.

**Event Component Inclusion**

Component inclusion was scored 1 or 0 for Manner, Path, and Goal in each description. A score of 1 indicated that a description included a certain component, a score of 0 indicated that the description did not include the component. Positive inclusion for each component required that certain criteria were met. Explanations for each event component’s criteria follow.

Manner was counted in either language if the description provided information about any of the Manner qualities outlined by Slobin (2006): motor pattern, rate of motion, force dynamics, attitude, or instrument. In English, a Satellite-Framed language, this is prototypically accomplished through main verbs (*hop, slide*), but Manner may also be encoded by other constructions. These alternatives include adjuncts, such as *by hopping*, or *in a way that resembled hopping*, adverbs describing the motion (*moved slowly*), and other elaborations about the body position of the animal such as (*moved on its back*). Verbs that only provided non-specific information about the Manner of the directed motion did not count for a positive Manner score. These verbs included *go, got into, moved, or traveled*, and other verbs that expressed that motion occurred, but not information about any of the Manner qualities outlined above. In Japanese, Manner may
also be expressed in the main verb such as \textit{aruku}, (walk), “[I] walk,” but as a Verb-Framed language it is more commonly expressed outside of the main verb, reserving that position for Path expressions. In these cases Manner may be encoded in a preceding subordinate verb such as \textit{hashi-tte iku}, (run-connective go), “[I] run,” or in the first, fixed morpheme of verb compounds such as \textit{tobi-agaru}, (jump-ascend), “[I] jump up.” Manner can occur even farther from the main verb adjuncts such as \textit{aruku yoo-ni}, (walk way-coordinative), “in a way like walking,” or in mimetics – typically reduplicated, onomatopoetic Manner expressions that may either stand alone (e.g., \textit{guruguru} “rolling”) or be marked with a coordinating particle \textit{to} (e.g. \textit{guruguru-to}). The forms with and without a particle are not distinct in their meanings.

Path of motion was considered included in either language if trajectory of motion was mentioned. In English this is prototypically achieved using satellites to the main verb, such as directional particles (\textit{up, down, over, backwards, away}), or prepositions (\textit{up, down, over, toward, along}). Other possible Path satellites include some adverbs (\textit{straightforwardly}) and phrases (\textit{in a straight line}). Path may also be encoded in main verbs in English (\textit{ascend, descend, climb, fall, process, back up}). In Japanese, Path expressed in the main verb is the most prototypical Path expression, such as \textit{agaru} (ascend), “[I] ascend.” Path may also be encoded in marked noun phrases such as \textit{ue-ni}, (up-locative), “up,” or \textit{migi-gawa-ni}, (right-direction-LOC), or with certain postpositional particles. The particles \textit{made, to, until,” and e}, “to” include Path information, while the particle \textit{ni}, “to,” does not (Beavers, 2008). This distinction is similar to the one between
over to and to in English. The Japanese locative particle *ni* and the English preposition *to* are Goal markers, and do not specify the Path of the event.

Goals of events were counted as included in either language if specific mention was made of the object or location at which the motion ended. This was achieved in English by using a prepositional phrase (*to the house, toward the tree, into the bowl*). Simple locative expressions such as *there* we also counted for Goal inclusion, if they clearly referred to the Goal. In Japanese, essentially the same criterion applied. Positive Goal inclusion was earned with the locative particles *ni, e,* or *made,* all meaning “to,” followed by a noun phrase, such as *ki-ni,* (tree-LOC), “to the tree,” or mention of an action at the Goal object earned Goal inclusion. As in English, simple locative expressions such as *soko,* "there," also counted for Goal inclusion.

Several examples of the coding are provided below in Table 3. Three English descriptions and three Japanese descriptions are presented along with their inclusion scores. Note that in general it was technically possible for participants to form a description of the experimental events including all components or none of them.

<table>
<thead>
<tr>
<th>Description</th>
<th>Manner</th>
<th>Path</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>A strange toy was moved across a table.</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Someone made the toy bunny slide down the ramp &amp; had it hop into the yellow bowl to the left of the ramp.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>The bunny landed in a cup</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Totoro-no ningyoo-ga ushiro muki-ni hanete, ichidan takai tokoro-ni noboru.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>&quot;The Totoro doll bounces facing backward and goes up one step higher.&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totoro-ga tanoshisoo-ni janpu shi-nagara ouchi-no mae ni itta.</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>&quot;Totoro went to the front of the house while jumping joyously.&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totoro-ga senaka-de subette ita.</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>&quot;Totoro was sliding with his back.&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Coded English and Japanese descriptions and their inclusion scores. Manner expressions are highlighted in orange, Path expressions in purple, and Goal expressions in green.

Inclusion scores for the three event components for each description were averaged per subject.
Results

Two omnibus ANOVAs were run, one analyzing effects across subjects and the other analyzing across items, which were the movies. The items analysis was necessary to examine inclusion across salience configurations. The subjects analysis ANOVA used language (Japanese vs. English), addressee (adult vs. child), and visual co-presence (present or absent) as between subjects factors, event component (Manner, Path, and Goal) as a three-level within subjects variable, and average inclusion scores across subjects as dependent variables. All significant effects and interactions from the subject analysis are shown in Table 4. Explanations for these effects are presented in the following 3 sections: Language Effects, Discourse Effects, and Salience Effects.

The items analysis ANOVA treated each movie as a different subject. There were 7 types of movies that varied what event component was most salient (Manner only, Path only, Manner and Path, etc.; see Table 1 for full descriptions of movies). These different salience configurations were the between subjects factor, and language (Japanese vs. English), addressee (adult vs. child), visual co-presence (present vs. absent), and component were within subjects variables; average inclusion scores was the dependent variables. All significant effects and interactions from the items analysis are shown in Table 5. List was also examined in these analyses, but because no main effect of list was found and it was a counterbalancing tool, we did not include it in subsequent analyses.
Table 4: Results of the subjects analysis ANOVA. Significant findings are shaded in gray.

<table>
<thead>
<tr>
<th>df</th>
<th>error df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Language</td>
<td>1</td>
<td>140</td>
</tr>
<tr>
<td>2</td>
<td>Addressee</td>
<td>1</td>
<td>140</td>
</tr>
<tr>
<td>3</td>
<td>Co-presence</td>
<td>1</td>
<td>140</td>
</tr>
<tr>
<td>4</td>
<td>Component</td>
<td>2</td>
<td>139</td>
</tr>
<tr>
<td>5</td>
<td>Language x addressee</td>
<td>1</td>
<td>140</td>
</tr>
<tr>
<td>6</td>
<td>Language x co-presence</td>
<td>1</td>
<td>140</td>
</tr>
<tr>
<td>7</td>
<td>Language x component</td>
<td>2</td>
<td>139</td>
</tr>
<tr>
<td>8</td>
<td>Addressee x co-presence</td>
<td>1</td>
<td>140</td>
</tr>
<tr>
<td>9</td>
<td>Addressee x component</td>
<td>2</td>
<td>139</td>
</tr>
<tr>
<td>10</td>
<td>Co-presence x component</td>
<td>2</td>
<td>139</td>
</tr>
<tr>
<td>11</td>
<td>Language x addressee x co-presence</td>
<td>1</td>
<td>140</td>
</tr>
<tr>
<td>12</td>
<td>Language x addressee x component</td>
<td>2</td>
<td>139</td>
</tr>
<tr>
<td>13</td>
<td>Language x co-presence x component</td>
<td>2</td>
<td>139</td>
</tr>
<tr>
<td>14</td>
<td>Addressee x co-presence x component</td>
<td>2</td>
<td>139</td>
</tr>
<tr>
<td>15</td>
<td>Language x addressee x co-presence x component</td>
<td>2</td>
<td>139</td>
</tr>
<tr>
<td>16</td>
<td>Language x addressee x x component</td>
<td>2</td>
<td>139</td>
</tr>
<tr>
<td>17</td>
<td>Language x addressee x component</td>
<td>2</td>
<td>139</td>
</tr>
<tr>
<td>18</td>
<td>Language x addressee x salience</td>
<td>6</td>
<td>8</td>
</tr>
</tbody>
</table>
Table 5: Results of the items analysis ANOVA. Significant findings are shaded in gray.

Language Effects

Before examining the inclusion scores in the omnibus ANOVAs, analyses were first run to determine whether speakers of English and Japanese differed in the quantity of speech in their descriptions, as measured by the average number of sentences per description. Using a univariate ANOVA with language (English vs. Japanese), addressee (adult vs. child), and visual co-presence (present vs. absent) as factors and average number of sentences as the dependent variable, a main effect of language was not found (F(2, 139) = .069, p = n.s.). Other effects were found in this univariate analysis, but they will be addressed in later sections. Furthermore, language interacted with neither addressee (F(1, 140) = .031, p = n.s.), nor visual co-presence (F(1, 140) = .246, p = n.s.). This suggests that whether speakers spoke Japanese or English did not affect how many sentences they used in their descriptions.
We then turned to the average inclusion scores of participants, used in the subjects analysis ANOVA displayed in Table 4. This analysis revealed a main effect of language (row 1) and this finding also contributed to an interaction between language and component (row 7). Japanese speakers describe Path more often than English speakers (mean Path inclusion score .85 > .70). This finding not only aligns with our typological expectations of Japanese as a satellite-framed language, but also supports the Whorfian prediction that one’s language has a direct influence on what speakers find most important to describe. The interaction was further supported by a series of post-hoc one-way ANOVAs comparing Manner, Path, and Goal inclusion across the two language groups. Path inclusion was different between languages (F(1) = 29.589, p < .001), but Manner (F(1) = .498, p = n.s.) and Goal (F(1) = .016, p = n.s.) were not. Component inclusion data are shown in Figure 1.

Figure 1: Average component inclusion scores for English and Japanese speakers.
The language main effect also informed many interactions shown in the subjects analysis and the items analysis. A three-way interaction between language, co-presence, and component was also found in the subject analysis (row 13), suggesting different patterns for handling co-presence of addressees in either language. This finding is the consequence of an independent co-presence effect, and will be further discussed in the Discourse Effects section. As shown in the items analysis in Table 5, language also interacted with addressee (rows 18, 26), salience configuration (rows 9, 15, 18, 21, 30), and component (rows 8, 19, 22, 26, 30). Further examination of the data revealed that these interactions were the product of the language main effect and other independent effects that will be discussed in following sections. Together, the language findings suggest that the language groups described events differently, but only with respect to which components they included in their descriptions, not the length of their descriptions. Japanese speakers talked more about Path.

Discourse Effects

As with language effects, the first dependent variable we examined was average number of sentences per description in the same univariate ANOVA, with addressee and visual co-presence as factors. This analysis found main effects of addressee (F(1, 140) = 4.944, p = .028) and visual co-presence (F(1, 140) = 9.918, p = .002), and an interaction between the two (F(1, 140) = 6.183, p = .014). Speakers of both languages used more sentences in their descriptions when speaking to adults (1.42 > 1.20), and when
addressees were absent (1.48 > 1.15). As is shown in Figure 2, this interaction is driven primarily by participants describing to absent adults, who used more sentences than participants in other discourse conditions. The trend was also shown in the present child vs. absent child conditions, but not as robustly. This suggests that participants were sensitive to the needs of their addressees with respect to the length of their descriptions.

![Figure 2: Average number of sentences for participants in the present adult, absent adult, present child, and absent child discourse conditions.](image)

Next we analyzed component inclusion scores, using the subjects analysis ANOVA. As shown in Table 4, an interaction between addressee and component was found (row 9). This difference is shown in Figure 3. Participants who described the movies for an adult included Goal information more than those describing to a child (.72
> .63). This suggests that speakers of both languages prioritized different components based on the identity of their addressee.

A main effect of co-presence (row 3) was also found. Speakers of both languages described more Path and Goal information to addressees who were not visually co-present. This finding also contributed to a three-way interaction between language, co-presence, and component (row 13), shown in Figures 4a and 4b. Figure 4a shows component inclusion scores in co-present and absent conditions for English, and 4b shows the same data in Japanese. Examination of the graph revealed that the interaction came about due to language-specific patterns of inclusion based on shared visual co-presence. English participants described more Path and Goal but less Manner to absent
addressees, while the Japanese participants described more of all components to absent addressees.

Figure 4: Average component inclusion scores for English participants (4a) and Japanese participants (4b) in the visually co-present and absent addressee conditions.
Further discourse interactions were found, but they are the product of previously discussed independent findings. These interactions are shown in the items analysis in Table 5: addressee and co-presence interacted with language (rows 18, 26), component (rows 21, 22, 26, 30), and salience configuration (rows 18, 21, 30).

To recapitulate, discourse factors had an effect of both the length and the component inclusion of participants’ descriptions. Participants provided longer descriptions to absent adults and expressed Goal more often to adults. Patterns of component inclusion based on co-presence also differed between languages. In English more Path and Goal and less Manner were included to absent addressees. In Japanese more Manner and Goal but the same amount of Path were included to absent addressees.

Salience Effects

We first examined effects of salience on description length. In a new repeated measures ANOVA similar to the aforementioned items analysis, we used the 7 different salience configurations as a between subjects independent variable, language (English vs. Japanese), addressee (adult vs. child), and co-presence (present vs. absent) as within subjects independent variables, and average number of sentences as the dependent variable. No main effect of salience was found ($F(6, 8) = .751, p = n.s.$), and this analysis reflected the findings seen in the subjects analysis using number of sentences. We therefore determined that this analysis did not contribute further to our understanding.
To examine salience effects we turned primarily to the items analysis ANOVA shown in Table 5. First of all, independent of salience influences, a main effect of component was found (row 4): speakers were overall less likely to describe Goal (mean inclusion score .68) than Manner (.72) or Path (.78). Post hoc paired t-tests that in fact all component inclusion scores were different from one another: Manner and Path (t(140) = -3.105, p = .002), Manner and Goal (t(140) = 2.213, p = .029), and Path and Goal (t(140) = 5.018, p < .001). This is a surprising finding, since we presented participants with movies in which all event components were designed to be roughly equally salient, and we did not expect any component to be described more or less overall. This effect was likely the product of the previously seen discourse interaction seen in Figure 3, where participants describing to children described Goal less often than those describing to adults.

Recall that we made a strong prediction about the influence of salience configuration on what components are described in different movie types. Descriptions of movies with salience configurations containing highly salient event components should express those components. This prediction was supported by a main effect of salience (row 5), and an interaction between salience and component (row 15). The main effect simply indicates that inclusion differed for different salience configurations. The interaction indicates that components were described more when they were highly salient in the stimulus movie. Furthermore, there was an interaction between language, salience, and component (row 21). Japanese speakers, who overall described more Path than
English speakers, described Path almost at ceiling even when it was not highly salient (mean inclusion score .80). When Path was highly salient Japanese described Path at even higher rates (.88). For English speakers, when Path was not highly salient they only described it about half of the time (.50), and when it was they described it much more (.80, a .30 increase), but still less than their Japanese counterparts. Other components did not show a language interaction. Manner was described more by both language groups when it was highly salient than not highly salient (.85 > .55), and Goal followed this same pattern (.92 > .34). Participants of both languages were indeed sensitive to the manipulated levels of salience in the movie clips, and this is displayed in Figure 5.

![Figure 5: Average component inclusion scores for English and Japanese speakers when the salience of each component was highlighted and when it was not.](image-url)
Summary

Lengths of descriptions were not different by language group or salience configuration, but were affected by discourse conditions. Speakers used more sentences to describe events to adults, and to addressees who were not-co-present. For component inclusion, although many significant findings were found in both the subjects and the items analysis, careful inspection of the data revealed four independent findings that essentially explained all the results. Firstly, as predicted Japanese speakers described more Path than English speakers. Secondly, speakers described more Goal in the adult addressee condition than the child addressee condition. Thirdly, speakers of English and Japanese showed different patterns based on the co-presence of their addressees. English speakers described more Path and more Goal, and less Manner to absent addressees, and Japanese speakers described more Manner and more Goal to absent addressees. And finally, high salience components in movies were described more often than low salience components, as predicted.

Discussion

This thesis set out to compare the weight of language, discourse, and salience factors on event descriptions. Effects from all of these factors were borne out, as well as some interesting interactions. While some of our findings were expected, others were surprising, and we will attempt to account for these. We expected that lengths of
descriptions would differ according to discourse factors (addressee identity and visual co-presence) but not by language or salience configuration, and this expectation was borne out. Speakers used more sentences when speaking to adults who were absent than other discourse conditions (Figure 2). Even though our experiment was not a cooperative audience design of the kind used by Lockridge and Brennan (2002), meaning that participants were not actually engaging with other speakers, they still made use of the needs of hypothetical addressees to plan the level of complexity of their descriptions. Furthermore, the discourse manipulation was between speakers, so this finding was not the result of participants adjusting their descriptions for new speakers or new discourse conditions. We can conclude that participants were at least sensitive to our discourse manipulations, but since they were so telegraphed at the onset of the experiment, it is unclear whether such addressee and co-presence factors would yield the same effects in a natural language setting. An experiment conducted in a lab setting where participants describe to different real addressees would more closely model a natural language situation. Also, early judgments based on discourse factors can be reanalyzed and changed later on in the discourse. That is, speakers revise their early conceptions in the presence of new information, and descriptions made later on in an interactive discourse may be different from those produced early on. The current investigation has very little to say about the development of discourse representations; however, the fact that participants created different length descriptions outside of a cooperative exchange suggests that such discourse considerations are taken into account even when other
knowledge about common ground is unavailable to the describer.

Our use of sentences as a measure for syntactic quantity of descriptions is somewhat crude. One benefit of using a cruder measure of language is that finer quantitative measures used in past work such as syntactic units or intonational phrases differ more dramatically across languages than a course measure, and therefore are almost incomparable (Barr, 2008). This is problematic for a cross-linguistic account of descriptions. If properly motivated, however, an analysis of the structure within sentences would be very valuable. For example, our analysis did not investigate how event components were ordered by speakers, and how densely the components were packaged. Allen et al. (2007) offered a potential analysis of clause structure of event descriptions in English, Japanese, and Turkish. They presented 40 speakers of each language, half of whom were adults and half were children, with animated DM events, and recorded and transcribed their descriptions. Following Talmy's typology of different languages' characterizations of motion events, they analyzed the syntactic packaging of Manner and Path within clauses of descriptions and coded how densely packaged components were within clauses as Tight, Semi-Tight, or Loose. These categories were only very loosely defined in terms of examples in all three of the languages, and did not appear to be consistent even within one typological language group. That said, the categories for syntactic packaging that they used can be paraphrased as follows: Tight packaging referred to Manner and Path being encoded by adjacent elements within the same clause. Semi-Tight packaging referred to when Manner and Path are encoded in
more distant elements, still within the same clause. Finally, Loose packaging referred to Manner and Path encoded in different clauses.

Their expectation was that, given that the Path is the core of a DM event, languages should keep Path in as prominent a position as possible, meaning as close to the main verb as possible. Manner on the other hand can freely be encoded further from the verb, as it is not as crucial to the DM event. Therefore, Satellite-Framed languages should package Manner and Path more Tightly than Satellite-Framed languages. They demonstrated the expected finding, but this is problematic. As previously mentioned, their categories of syntactic packaging are poorly defined, and seemingly inconsistent within languages. For example, two of the Japanese examples for Tight and Semi-Tight given only differ in the presence of the coordinating particle to, which should not change the meaning of the utterance.

This example highlights another problem. An account of syntactic packaging should appeal to the syntactic distance of components, that is, how close they are compositionally in the phrase structure. If Allen et al’s framework did provide such a formal basis for syntactic closeness, then the existence of a coordinating particle alone should not constitute a different level of Tightness for syntactic packaging. Intuitively, their notions about how event components are encoded together within clauses seem potentially interesting, but their model of what this might look like is incomplete. Future work is needed to see whether Verb and Satellite-Framed languages package event components differently within clauses.
In our study, as expected, all components were described more often when their presentation was highly salient in the stimulus event (Figure 5). This finding supports our claim that more salient components, that is components that are more accessible or imageable from the observed, gain higher priority to be included by a speaker in a description. Participants overall described Goal less often than Manner or Path, but this was due to an addressee interaction: participants described less Goal information to children than adults (Figure 3). This could have several potential explanations. Recall from our introduction of DM event component salience that Goal salience is somewhat less nuanced than either Manner or Path salience. Goal salience is relative mainly to the existence of an object at the location where the figure ends its motion. In low salience conditions, the object simply was not there. For Manner and Path low salience conditions, on the other hand, the properties of the Directed Motion were still available (in fact, they had to be, otherwise they would not have been DM events), but they were made less prominent or perhaps less noticeable. Overall, we expected all three components to have the same relative salience both in the high and low salience configurations, but examination of the low salience bars in Figure 5 suggests that perhaps low salience Goal was less salient than either Manner or Path. If that is was case, if adults were trying to make their descriptions simple for children by omitting unnecessary information, Goal would seem to be a good candidate for omission. An alternative explanation for the addressee interaction is that the perceived aspectual objectives of describing to adults vs. children were different. That is, perhaps participants describing
to children talked more about the properties of the motion without feeling obligated to talk about the endstate of the event, while participants describing to adults deemed all event components equally important.

Another explanation then for the Goal disparity is that Goal as a component is more readily omitted by participants when it is in low salience. Consider for a moment the other two components. Manner is composite of many different qualities of motion, providing participants with more opportunities to describe Manner. Since describing any one of these qualities or more earned a participant positive Manner inclusion scores, it is not surprising that Figure 5 shows that Manner is described more than Goal in low salience. For Path, the trajectory of the motion, while we manipulated the movies so that in some movies the Path had a vertical component (on the ramp) and others did not (on a flat board), in either case a clear trajectory existed. However, Goal salience hinged solely on the presence of an object at the location where the Figure stopped moving. If an object was present, we considered it high salience, and if an object was absent, we considered it low salience within the frame of the experimental events. Since the distinction is simpler for Goal than or Manner or Path, it is reasonable that our participants would show more definite behavior when Goal is either in high or low salience; that is, describe is more when it is highly salient and less when it is in low salience than other components. It was perhaps not described more than other components because all three were described at ceiling levels when they were highly salient, but Figure 5 shows that Goal was described less than other components when it
was in low salience.

We expected speakers of Verb-Framed and Satellite-Framed languages to conform to the typological expectations about Manner and Path encoding. Specifically, we expected Manner to be described more by English speakers and Path to be described more by Japanese speakers. The latter prediction was attested by the data, but not the former. Japanese participants described Path at near ceiling levels even when it was in low salience (Figure 1). At first glance, this finding suggests a Whorfian influence from language onto thought for Japanese speakers: their perception of the events was influenced by their language’s prototypical structuring of events. However, let us consider some facts about Japanese and English. There are more Path verbs in Japanese, and verbs are obligatorily encoded in a clause. In addition, these Path verbs (e.g. *agaru*, "ascend," *susumu*, "proceed," etc.) obtain higher frequency than their English counterparts. Still other Path verbs in Japanese (e.g. *iku* "go," *idoosuru* "move," etc.) have English equivalents that do not express Path information. Therefore, it is simply easier to express Path in Japanese. Beavers (2008) pointed out that in order to express an overt Goal, a Manner verb cannot be used on its own in Japanese, a Path verb must be taken as the main verb in the clause. Otherwise, with a Manner verb alone, a Directed Motion interpretation cannot be obtained from the utterance. In English, where Manner is prototypically encoded in the main verb and is the component that was typologically predicted to be described more, there are verb options available to speakers that do not encode Manner (e.g. *move, go*), perhaps more options than those verbs available to
Japanese speakers that do not include Path information. In other words, perhaps in Japanese it is difficult *not* to express Path. These facts suggest another explanation for the language effect. It is not necessarily the case that speakers attended to different components of the event, simply that the linguistic systems favored different components.

Our co-presence finding is puzzling (Figure 4). To break it down, first recall that Japanese participants described more of all components to absent addressees than co-present addressees. This finding by itself was unexpected but understandable. If participants had less shared information with absent imagined addressees, they may have felt that they needed to include more information to absent addressees in order to be as informative as possible. Moreover, participants may have omitted information that seemed inferable to co-present addressees who also saw the event. English participants showed this trend as well, except for their Manner inclusion. They described more Manner to absent addressees. It is odd that, since Manner is the typologically preferred component for English speakers, it would be described less often in a discourse condition in which they shared less information with their addressees. One possible explanation for this finding is that Manner, which is the most complex of the components whose inclusion we measured in this task, was deemed by English speakers to be the most difficult to convey to speakers who did not view the event, and so omitted it. This explanation is far from satisfactory, as it is very common to describe the Manners of events to people who did not view the event. Furthermore, this does not explain why Japanese speakers did not make a similar judgment. This finding highlights the need for
a finer-grained measure of component inclusion. Since inclusion scores in our study were binary per description, we did not examine multiple mentions of components or their ordering within descriptions. Our analysis also did not take into account the syntactic structures used to encode components, which would have typological predictions and potentially interesting interactions with ordering and syntactic packaging (Allen et al., 2007), as well as discourse conditions.

Our study found that while description length was only affected by discourse conditions, component inclusion was affected by language, discourse, and salience factors. We also found that our discourse conditions were taken into account by participants even though they were not taking part in an interactive exchange, and our salience manipulations led to expected greater or lesser component inclusion. We have yet to truly see how these influences weigh in against each other to inform descriptions and what their interaction is like online, but this study demonstrated that speakers are sensitive to factors of language, discourse, and salience when conveying information about motion events in descriptions.
References


57
