Language Processing and Disorders

Language Processing and Disorders

Edited by

Linda Escobar, Vicenç Torrens and Teresa Parodi

Cambridge Scholars Publishing

Language Processing and Disorders

Edited by Linda Escobar, Vicenç Torrens and Teresa Parodi

This book first published 2017

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-4438-9508-3
ISBN (13): 978-1-4438-9508-8

TABLE OF CONTENTS

A Review of Language Processing, Second Language Acquisition and Language Disorders 1
Vicenç Torrens, Linda Escobar and Teresa Parodi

Part One: Language Processing

Reaction Time as a Measure of Implicit Grammaticality Judgment..... 12
Misha Becker

Probabilistic Phonotactics in Vocabulary Acquisition during Reading in a Native Language 31
Denisa Bordag, Maria Rogahn, Amit Kirschenbaum and Erwin Tschirner

The Construal Hypothesis and Relative Clause Processing: The Effect of the Referentiality Principle in Brazilian Portuguese 54
Gitanna Brito Bezerra and Márcio Martins Leitão

Language Experience and Memory Effects in Anaphora Resolution in Greek 75
Eleni Fleva, Georgia Fotiadou, Maria Katsiperi, Eleni Peristeri, Maria Mastropavlou and Ianthi Maria Tsimpli

The Form-Function Relation in *of*-phrases: An Experimental Approach..... 92
Takashi Fujiwara, Fuminori Nakamura and Daisuke Suzuki

An Acceptability Study of Long-Distance Extractions in Swedish..... 103
Anna-Lena Wiklund, Fredrik Heinat, Eva Klingvall and Damon Tutunjian

The Cue-based Retrieval Theory of Sentence Comprehension: New Findings and New Challenges..... 121
Dan Parker, Michael Shvartsman and Julie van Dyke

Contextual Information in Universal-Q processing and Some Remarks on Distributivity, Collectivity and Maximality..... 145
Erica dos Santos Rodrigues and Mercedes Marcilese

Part Two: Second Language Acquisition and Bilingualism

When the Burden of Age Does Not Wait: Early and Late L2 Acquisition of Differential Object Marking in Spanish 166
Pedro Guijarro-Fuentes and Acrisio Pires

Retrieving Presupposition in English L2: An Eye-tracking Study of Pragmatic Scales with Focus Particles 193
Olga Ivanova

Small Clauses in the Bilingual Syntactic Priming Paradigm 213
Paulina Łęska and Katarzyna Jankowiak

Part Three: Language Impairments

Is the Self-reference of Autistic Children Atypical? The Case of Two French Autistic Children 236
Camelia-Mihaela Dascalu

Psycholinguistics of Dementia: What is Known about Specific Language and Speech Traits of Alzheimer's Disease? 258
Olga Ivanova, Juan José García Meilán, Francisco Martínez Sánchez, Emiliano Rodríguez and Juan Carro

Light Verbs Revisited: A Comparative Perspective of Impaired Language Development and Dementia 277
Vasiliki Koukoulioti and Stavroula Stavrakaki

Syntactic Complexity in Children with Autism Spectrum Disorder and Specific Language Impairment 291
Alexandrina Martins, Ana Lúcia Santos and Inês Duarte

Regular and Irregular Inflectional Morphology in Acquired Language Disorders: The Case of German 314
Martina Penke and Eva Wimmer

Pragmatic Coherence in Alzheimer's Disease: A Protocol for Analysis 345
Ana Varela

A Case of Surface Dyslexia in a Spanish-English Bilingual Patient with Broca's Aphasia 374
Cristina Vereda, Mercedes González-Sánchez and Lidia Taillefer

A REVIEW OF LANGUAGE PROCESSING, SECOND LANGUAGE ACQUISITION AND LANGUAGE DISORDERS

VICENÇ TORRENS, LINDA ESCOBAR AND TERESA PARODI[1]

Language processing is considered an important part of cognition and studies conducted in this field are increasing. This book gathers together a collection of papers on language processing, second language acquisition and language disorders. The first paper by Misha Becker, "Reaction Time as a Measure of Implicit Grammaticality Judgment" deals with children's acquisition of the argument structure of novel predicates. This study is based on a reaction time methodology, where children demonstrate a longer period of time to process an ungrammatical sentence or garden path. It presents an experiment in which children aged 4-7 years old had to answer a grammatical or ungrammatical question. The sentences included in the experiment differed with respect to correct or incorrect argument structure. The sentences had familiar and novel verbs, including transitive and intransitive verbs. Becker found that children usually had slower reaction times with ungrammatical sentences than with grammatical ones. In a second experiment, she tested younger children aged 3-4 years old. This time they had to answer questions that included transitive and intransitive familiar verbs. In this second experiment, she found that children answered the ungrammatical questions more slowly than the older children in the first experiment. These data suggest that children aged 3-7 years can distinguish between grammatical and ungrammatical questions in terms of the argument structure of the included verbs.

[1] Vicenç Torrens [corresponding autor], Facultad de Psicologia, U.N.E.D. E-mail: vtorrens@psi.uned.es
Linda Escobar, Facultad de Filologia, U.N.E.D. E-mail: maescobar@flog.uned.es
Teresa Parodi, Department of Theoretical and Applied Linguistics, University of Cambridge. E-mail: tp209@cam.ac.uk

The paper, "Probabilistic phonotactics in vocabulary acquisition during reading in a native language" by Bordag et al. investigates the role of the prominence of a word form in the incidental acquisition of new words in native German speakers. Specifically, they studied the role of the relative frequencies of segments and their sequences in words. In order to achieve this goal, the authors collected data on the inference and establishment of meaning, its integration within the existing semantic network and the establishment of a new word form. In a self-paced reading task, the authors found that participants learned and retained, in memory, the meaning of both high and low salient novel words, although the reaction times were longer in the implausible condition than in the plausible condition. In a second task, with a vocabulary knowledge scale, participants had to judge their knowledge of novel words they had learned in the previous task, along with totally novel words. The authors found that participants could more easily recall a word with a low phonotactic probability than a word with a high phonotactic probability. In a third experiment, using a semantically-primed lexical decision task, the authors found that participants more slowly processed targets with a semantically-related novel word prime than an unrelated existing word. This semantic effect is totally different for words with firmly established representations. In a fourth experiment, with a recognition task with repetition priming, the authors found no differences in priming effects between high phonotactic probability and phonotactic probability low novel words. They concluded that phonotactic probability in novel words is crucial to understanding lexical acquisition, which varies depending on different variables.

The aim of Bezerra & Leitão's study was to conduct experimental research on the referentiality principle and, in more general terms, on the predictions that the construal hypothesis makes with respect to Brazilian Portuguese. In particular, this study addresses the referential status of the N2 and its effects for relative clause processing and the initial syntactic under-specification proposed by the construal hypothesis. Given the fact that the referentiality principle has been primarily studied through off-line experiments, this method used on-line and off-line experimental techniques to provide new experimental data through: (i) a self-paced reading task; (ii) a questionnaire.

Other researchers are also interested in complex phenomena such as anaphora resolution (AR). In the paper, "Language experience and memory effects in anaphora resolution in Greek", by Eleni Fleva, Georgia Fotiadou, Maria Katsiperi, Eleni Peristeri, Maria Mastropavlou and I. Maria Tsimpli, AR is examined by contrasting more-or-less complex sentences that include overt or null pronouns. In their experiment,

participants were presented with two self-paced listening (SPL) tasks under different conditions. However, in both tasks, anaphora resolution of the null and overt pronoun was assessed with subject-verb-object (SVO) sentences and with sentences that presented a rather different word order, including clitic-left-dislocation (O-cl-V-S) structure. A number of factors were taken into account such as preference for the object or the subject-referent of the sentence, as well as the time needed for resolution. The authors discuss interesting differences in participants' performance, taking age, education, language experience and memory resources into account, among other things.

Takashi Fujiwara, Fuminori Nakamura and Daisuke Suzuki deal with the issue of synonymy in their paper, "The form-function relation in of-phrases: An experimental approach", to discover whether "of-phrases" and adjectivals (i.e. "of interest vs. "interesting") are processed equally in English. They show that the form-function relation of "of-phrases" plays a significant role in the processing of English prepositional phrases. Considering previous corpus analysis, they select two factors for their analysis: "subject type" (how the subject slot is filled) and "modality" (whether or not the expressions co-occur with modal verbs). In this way, they could plan their experimental study through a questionnaire, paying more attention to these two variables in unison rather than individually. Traditionally, "of-phrases" are grammatically equivalent to adjectivals. Nevertheless, there is no fine-grained description of the differences between the two. The experimental findings reported in this experimental study focus on the previous highly subjective variables and how discourse function plays a role, providing a significant way for future comparison of these two expressions.

The purpose of the study by Wiklund et al. was to obtain controlled acceptability judgment data for Swedish (but also probably extended to other mainland Scandinavian languages) regarding structures that have been assumed to not involve island-like violations, like in most other languages. In particular, three structures in their extracted and non-extracted forms are under discussion: relative clauses (RC), that-clauses (TC), and non-restrictive relative clauses (NRC). Contrary to what should be expected, the results obtained did not always align with the on-line measures obtained via eye-tracking in other previous studies reported in the literature.

Parker, Shvartsman and Van Dyke's paper provides an account of current perspectives on memory retrieval in sentence comprehension. After discussing many psycholinguistic results on the timing and accuracy of dependency-formation found in a large number of present experimental

studies, which are carefully reviewed, they argued that these findings can be best captured with "a direct-access retrieval mechanism that gives preferential weighting to syntactic information when navigating linguistic representations in memory."

Inspired by the work of Lima (2013), but using a slightly different methodology, Erica dos Santos Rodrigues and Mercedes Marcilese present new findings and results from two experiments designed to examine the possible role of preceding context in the interpretation of potentially ambiguous Q-expressions. In their study, "Contextual information in Universal-Q processing and some remarks on distributivity, collectivity and maximality", these authors investigate the processing of three Q-expressions, "cada", "todo", and "todos os", in Brazilian Portuguese (BP), with a larger number of participants than were in the previous research reported in the literature. They also present two further off-line experiments intended to research how the maximality property of universal quantifiers is associated with the processing of "todo" and "todos os" in BP. Interestingly, they discuss the differences found in the processing of these two apparently similar quantifiers where context plays a decisive role.

In their paper, "When the burden of age does not wait: early and late L2 acquisition of differential object marking in Spanish", Guijarro & Pires examine the acquisition of differential object marking (DOM) in L2 Spanish, by speakers of English exposed to Spanish before puberty and later, thus testing the critical period hypothesis. The results show that age is not a predictor for accuracy in the use of the Spanish preposition "a" in the relevant DOM contexts. The authors discuss the results against the background of the interpretability and feature reassembly hypotheses. According to the former, uninterpretable features are expected to be problematic for L2 learners, but not interpretable ones. According to the latter, the acquisitional problem is orthogonal to interpretability, as even with a similar set of L1 features, L2 learners may find difficulties in their target realization. The results appear to challenge the interpretability hypotheses: interpretable features turn out to be problematic for both early and late learners. On the other hand, the fact that semantic (interpretable) features are difficult to acquire is interpreted as evidence in favor of the feature reassembly hypothesis.

Olga Ivanova's paper, "Retrieving presupposition in English L2: an eye-tracking study of pragmatic scales with focus particles" deals with how L2 speakers of English with German L1 process scalar expressions with the focus particle "even". The study compared processing in L2 to native processing of focus particles which, as opposed to L2, has already been extensively studied. In a self-paced reading task, the test items were

contrasted according to the occurrence (or not) of "even" in the utterance "Mike and Lucy love (even) punk". The critical items measured by eye-tracking on first and second pass were "even" and the focus "punk". These time measures were taken to reflect processing cost. The outcome showed that in the utterance without a discourse particle there were significant differences in the low- and high-level responses between English natives and German L2 English speakers; this was taken to indicate that cultural background may play a role. On the other hand, the presence of the discourse particle "even" yields similar results for both groups of participants: in both cases, "even" involves the highest processing cost as measured by reading times, both in the first and the second pass. According to the author, this result indicates that focus particles are universal tools that guide pragmatic inference independent of proficiency.

The paper by Łęska & Jankowiak tries to find syntactic priming in English and Polish bilingual populations. They tested whether processing of a structure in one language was enhanced after using an analogue in another language. The structure under study was small clauses, present in both English and Polish. Small clauses are predication phrases with the semantic function of predication. Two types of small clause constructions were used: small clauses whose predicate is an adjective (AP predicates); and small clauses whose adjectival predicate is introduced by a preposition (PP predicates). This research compares the reaction times elicited by stimuli presented in L1 as compared to L2, as there are usually longer reaction times in a non-native (rather than a native) language. There were three experimental blocks: two within-language blocks (a within-language Polish, and a within-language English block); and one between-language block. The participants had to decide whether sentences were grammatically correct or incorrect. The authors found that accuracy rates were higher for primed compared to unprimed conditions in both languages, regardless of the small clause type. The authors found that sentences are easier to process when they are primed with the same syntactic structure in both languages.

"Is the self-reference of autistic children atypical? The case of two French autistic children" by Camelia-Mihaela Dascalu, addresses the question of whether non-typically-developing children have difficulty in learning pronominal reversals, on the assumption that there are a large number of causes that can produce this difficulty, such as echolalia, the mother's input, conversational roles, the understanding of context, or the lack of the theory of mind. Dascalu conducted research on the use of self-reference by autistic children. The findings were compared with those of a typically-developing child. The data from the autistic children were

collected at home with audio-visual recordings of spontaneous speech interaction with mothers and other children. The longitudinal corpus of one neuro-typical child was collected manually and transcribed with CLAN. Dascalu found that the typically-developing child produced linguistic markers very early on, whereas the autistic children rarely used their first names in subject position. Instead, they used the third and second person pronouns to refer to themselves. Dascalu argues that autistic children have a particular difficulty with perspective-taking and reference shift, which explains the delay in the development of the pronoun system in this population. She argues that this delay is not due to language development but a lack of cognitive flexibility.

"Psycholinguistics of Dementia: what is known on specific language and speech traits of Alzheimer's disease?" by Ivanova et at., is a review of studies on language impairment in dementia. The paper tries to supply a new method to measure a prosodic aspect of language. This method attempts to facilitate an early and efficient diagnosis of Alzheimer's disease (AD). These patients have lexico-semantic problems, due to the disruption of the networks of semantic knowledge. This causes a loss of concepts and other difficulties in lexical access. The authors argue that lexical-based assessment can predict the probability of the onset of AD, since lexical impairment usually causes anomic aphasia in the disease. In addition to this, looking into properties of intonation, shifts in stress or in the amount of sound are found along with voice breaks in prosody, or variations in rhythm or intonation, in patients with AD. Finally, these authors analyze the voices of people with non-pathological and pathological ageing, in order to identify and to extract the prosody-related acoustic and vocal parameters associated with the disease. Ivanova et al. found a high percentage of deaf segments, a significant difference in the prosodic parameters, a reduced speed of elocution and articulation in reading, a low effectiveness of phonation time, an increasing number and proportion of pauses, and significant changes in the speech analyses in the voices of AD patients, compared to non-pathological people.

Children with specific language impairment and individuals with different types of primary progressive aphasia have been found to overuse light verbs, which have an underspecified semantic representation and an incomplete argument structure. Koukoulioti & Stavrakaki report the results of a naming and a comprehension task administered to four children with SLI, and a study on a sentence elicitation task administered to seven patients with semantic dementia. They found that the two groups produced a high percentage of light verbs when trying to name actions and that semantic dementia patients relied more on light verbs than the SLI

children. The structure of the light verbs constructions was preserved in these patients but the semantic components were not as precise as in a normal population.

The paper by Martins, Santos and Duarte, "Syntactic complexity in children with Autism Spectrum Disorder (ASD) and Specific Language Impairment (SLI)", compares the performance of typically-developing (TD) children to those with SLI and ASD, in the production and comprehension of relative clauses in European Portuguese. The focus lies in subject and object relative clauses in simple clauses and in those with one further level of embedding, i.e. where the subject or object is extracted from an embedded relative clause. Production was tested by means of elicited imitation and the results showed that the SLI and ASD children displayed a lower performance than age-matched TD peers; in simple relatives the ASD children performed better than the SLI children and a subject-object asymmetry was not in evidence. A careful analysis of errors shed light on how the different groups dealt with the subject-object asymmetry and the level of embedding. Comprehension, in turn, was tested by a truth-value judgment task. The performance was similar to that of the production task, with some group-specific differences. In the TD group there was a more obvious subject-object asymmetry, as well as an effect in terms of the level of embedding for the younger children. As for the SLI and ASD groups, it appears that the level of embedding affected the ASD group less. These differences in error types are an interesting result and are potentially indicative of a production-comprehension contrast.

Penke and Wimmer's paper deals with explanations of the neural basis of language disorders like Broca's and Wernicke's aphasia and Parkinson's disease. The single mechanism approach assumes that regular and irregular inflected forms are generated by only one cognitive mechanism; the dualistic view says that regular and irregular morphology have different underlying processes. Following the dualistic view, only irregular forms are stored and retrieved as fully inflected forms from the mental lexicon. If two different components underlie regular and irregular inflection, language disorders might affect the regular or irregular inflectional processes, depending on the areas of the brain that have been damaged. Broca's aphasia and Parkinson's disease are expected to result in difficulties with regular inflection whereas patients with Wernicke's aphasia are assumed to have difficulties with irregular inflected forms. The authors of this paper found that in Wernicke's and Broca's aphasics and patients with Parkinson's disease, the error rates for irregular inflected participle and noun plural forms were higher than for regular participle

and noun plural forms. These findings in German differ from the predictions drawn from the current models on the representation and processing of inflection in the brain: Broca's aphasics and patients with Parkinson's disease were not selectively impaired in producing regular inflected forms and were even able to over-regularize regular inflection to irregular inflecting stems. The different error rates observed in English versus German were due to differences in the inflectional systems in these languages. These authors propose that the three groups of patients were suffering from processing difficulties.

The paper by Varela presents a model on coherence production and reception, which was tested by analyzing the speech of a patient with moderate dementia of the Alzheimer's type. The model or coherence proposes four steps for uttering coherently: (i) the intention of our statement; (ii) the cognitive factors that can influence the quality and appropriateness of our statement; (iii) the pragmatic factors to produce a meaning given the context; and (iv) linguistic factors so that we can produce and understand our utterance correctly. The understanding process follows the following route: (i) processing the stimulus with our cognitive abilities; (ii) considering the context we are surrounded by; (iii) decoding the message according to this context; (iv) getting the communicative intention. The patient under study showed many pragmatic deficits in her discourse, difficulties in naming and complex syntactical structures, difficulties with lexical retrieval and management of shared knowledge. Personal orientation deficits caused problems in processing the participants of the conversation and time orientation deficits were responsible for the lack of relevance. This model could be useful as a tool for identifying coherence deficits and developing speech therapies.

Vereda presents a case study of a patient suffering from acquired Broca's aphasia with acquired surface dyslexia. Taking into account that the subject was a Spanish dominant bilingual, they checked if acquired surface dyslexia was present in Spanish and English. In order to find out if there was any significant difference between both languages in terms of reading, they conducted lexical decision and reading tasks. This author found an extremely significant difference in spelling-sound regularity for exceptional words in English ($p < 0.01$). Significance was also found for regularly inflected words, showing an influence from Spanish pronunciation for regular inflected verbs in the past tense. This influence was also observed when words were similarly spelled in both languages.

The present volume deals with research on language processing and disorders presented at the Experimental Psycholinguistics Conference in Madrid. It covers topics ranging across syntax processing, second-language

acquisition, lexical processing and language disorders. We would like to thank the plenary speakers (Roelien Bastiaanse, Manuel Carreiras, Volker Dellwo, Teresa Parodi, Jeannette Schaeffer) and the members of the scientific and organizing committees (Joanna Blaszczak, Denisa Bordag, Carlo Cechetto, Albert Costa, Naama Friedmann, Anna Gavarró, Nina Jeanette Hofferberth, Olga Ivanova, Victoria Marrero, Silvia Martínez-Ferreiro, Eva Moreno, Robert Reichle, Jeannette Schaeffer, Eva Soroli, Ellen Thompson, Spyridoula Varlokosta and Monica Wagner). We are also very grateful to UNED for their support to organize this edition of the conference.

PART ONE:

LANGUAGE PROCESSING

REACTION TIME AS A MEASURE OF IMPLICIT GRAMMATICALITY JUDGMENT

MISHA BECKER[1]

Abstract

The purpose of this paper is to introduce an experimental methodology to measure children's implicit judgments of grammaticality or acceptability; that is, to assess their judgments of well-formedness without requiring them to supply a metalinguistic grammaticality judgment. The methodology uses Reaction Time (RT) to gauge a child's degree of "surprise" at hearing some linguistic form. The premise of this methodology is that it will take a speaker longer to respond to an ungrammatical prompt (in the implementation here, a yes/no question) than a grammatical one. Although RT is a familiar and long used methodology in studies of language processing and other domains of cognition, it has not been used previously in the way outlined below. This paper is structured as follows. In section 1, I present the rationale for using RT to measure grammaticality judgment in children: I explain why existing methodologies of assessing children's grammatical knowledge are insufficient and why RT succeeds where other methods fail. In section 2, I describe the methodology in detail, including how it is carried out and how the results can be analyzed. I also present some preliminary results obtained using this methodology. Finally, I address some open questions and directions that I hope this work will take in the future.

[1] Misha Becker, University of North Carolina, Chapel Hill. E-mail: mbecker@email.unc.edu.

These results have been published in Becker (2014, 2015a, b); thus the contribution of this particular paper is not the results themselves but a focus on how the studies were conducted.

1. Rationale

1.1. Grammaticality Judgment

Modern linguistic theory seeks to develop a formal model of the knowledge of a language that a speaker has in his or her mind, and to understand how that knowledge is structured and acquired. For many linguists, the real point of interest is the tacit body of knowledge known as "linguistic competence" (Chomsky, 1965). Linguistic competence is taken to be a quasi-pure form of our linguistic knowledge, unmarred by problems that can interfere with the actual usage of language ("performance"), including slips of the tongue and other errors of speech production and comprehension. In the study of adult grammatical knowledge, a common approach has been to rely on native speaker judgments about sentence well-formedness or acceptability (Chomsky, 1965; Newmeyer, 1983; Schütze, 1996; Cowart, 1997). As observed by Schütze (1996), there are a number of advantages of using this method to learn about linguistic knowledge over simply observing naturalistic speech production: native speaker judgments give us information about constructions that are grammatical but only rarely used in natural speech; they allow us to examine speakers' knowledge about ungrammatical forms; and, given that people do make speech errors in production, obtaining grammaticality judgments can help us to tease apart the difference between forms produced that actually conform to the speaker's mental grammar, from forms produced that may in fact be slips of the tongue.

However, this methodology has been criticized along several dimensions (Hiramatsu, 2000). A number of researchers have questioned the utility and accuracy of data collected via metalinguistic judgment, as a means of drawing inferences about speakers' underlying competence (Birdsong, 1989; Wasow & Arnold, 2005). Others have noted variability across speakers in the kinds of judgments given (Labov, 1975). Snyder (2000) cautions against the problem of syntactic satiation, whereby certain syntactic violations (notably *whether*-island violations and complex NP violations) begin to sound more acceptable after multiple presentations of these ungrammatical sentences (cf. Sprouse, 2009 for an alternative account of the satiation effects obtained by Snyder). And there is disagreement over how native speaker judgments should be elicited, i.e. through a simple binary judgment (OK/not OK), Magnitude Estimation (Bard et al., 1996; Featherston, 2004), the Likert Scale, or forced-choice between two sentences.

Concerns about this methodology are exacerbated when we consider using it with young children. Of primary concern is whether, and at what

age, children are capable of providing reliable metalinguistic grammaticality judgments. Metalinguistic awareness typically develops during the early school years as children learn other metalinguistic skills such as reading and writing (Gindes, 1980; Menyuk, 1981). A number of researchers have developed training protocols that can be used to successfully elicit grammaticality judgments in children ages 4 and older, with success occasionally extending to precocious 3-year-olds (McDaniel & Cairns, 1990, 1996). The training involves drawing children's attention to their own language and perhaps different languages they have been exposed to, and engaging them in discussion about how language works (e.g. that it has words, we use words in particular ways, etc.) and finally by demonstrating some ill-formed sentences and pointing out that these sentences "sound funny" or are said in "the wrong way". Using this type of procedure, researchers have successfully elicited children's grammaticality judgments concerning pronoun reference, subjacency, wh-movement, relative clauses and subject-auxiliary inversion (McDaniel, Cairns & Hsu, 1990; Stromswold, 1990; Maxfield & McDaniel, 1991; McDaniel, Chiu and Maxfield, 1995; McDaniel & McKee, 1995, among others). As I will argue, however, reliance on children's metalinguistic judgments is insufficient for studying particular areas of their developing grammars, notably their acquisition of abstract predicates and their argument structure properties.

1.2. Abstract Predicates

The area of grammar acquisition I am interested in is how children determine the argument structure properties of novel predicates. I am particularly interested in how children determine these properties for very abstract predicates, like *seem* or *easy*. With concrete predicates, such as *eat*, *run*, or *pink*, there is likely to be some experiential basis for determining at least the core aspects of the word's lexical semantics, and this can help constrain the likely argument structure requirements or syntactic properties of the predicate, roughly along the lines suggested by the Semantic Bootstrapping Hypothesis (Pinker, 1989). For example, if a learner can determine, through hearing the verb *run* used in conjunction with running events, that the verb *run* denotes a self-initiated motion, then the learner might expect this verb to occur with an Agent subject, and perhaps with a locative adjunct, but he or she would not expect it to occur with a sentential complement.

(1) John ran (to/in the park).
(2) *John ran that he was tired.

Conversely, as demonstrated by a wealth of studies in the Syntactic Bootstrapping literature (Gleitman, 1990; Gleitman et al., 2005, among many others), knowing the argument structure properties of a predicate can help narrow down the lexical semantics of the word. Thus, if a learner hears an unknown verb in the sentence frame in (3), he or she might assume that the verb denotes a motion but not a mental state. But upon hearing an unknown verb in the sentence frame in (4), the same learner is more likely to surmise that the verb denotes a mental state or a verb of communication, rather than a motion (Kako, 1998; Snedeker & Gleitman, 2004).

(3) John verbed (to/in the park).
(4) John verbed that he was tired.

Unlike more concrete predicates, however, for certain abstract predicates like *seem* and *easy*, neither of these sources of information is straightforwardly available to learners. Since these predicates denote, precisely, abstract states and properties, they are not directly observable through the environment, so that experiential sources of information about lexical meaning are compromised. More importantly, in my view, these predicates overlap in their surface distribution with other subclasses of predicates that have very different argument structure properties (see Becker, 2006; Becker & Estigarribia, 2013, for discussion of why this presents a learning puzzle). As is well known, raising verbs like *seem* occur in some of the same surface strings as control verbs, such as *claim* (Davies & Dubinsky, 2004).

(5) John seems [*t* to be friendly.]
(6) John claims [PRO to be friendly.]

In a parallel fashion, *tough*-adjectives, like *easy*, occur in some of the same sentence strings as control adjectives, like *eager* (Lees, 1960; Chomsky, 1964) (*t* indicates a trace of movement, while *e* indicates an empty position without movement).

(7) John is easy [PRO to please *t*]
(8) John is eager [PRO to please *e*]

Despite the surface similarity, these subclasses of predicates differ in important ways. Within movement-based approaches to syntax, the subject in (5) is claimed to undergo movement from the embedded subject position, while the subject in (6) is base-generated in the main clause and "controls" the reference of the embedded subject, PRO (Chomsky, 1980; Collins, 2005). Though somewhat more contentious, the surface subject of (7) has also been analyzed as raising into that position from a position in the embedded clause (Rosenbaum, 1967; Brody, 1993; Hicks, 2009), whereas, like (6), the surface subject of (8) is base-generated in the main clause.

While these analyses rely upon particular syntactic assumptions about movement and derivation, the argument structural differences between these predicates do not hinge on whether or not one's syntactic analysis involves movement. In terms of argument structure, raising verbs and *tough*-adjectives do not select an Agent or Experiencer subject argument, and therefore can occur with an expletive or inanimate subject, while control predicates (*claim, eager*) do select such an argument and therefore disallow expletives.

(9) It seems/*claims to be cloudy.
(10) It is easy/*eager to please John.

In previous work, I have discussed the theoretical issues surrounding the acquisition of these predicates (Becker, 2014, 2015a, b). In this paper, I will focus on a method for studying how children categorize them. That is, how does a child determine that a given predicate in their language has the properties of *seem* or *easy*, as opposed to *claim* or *eager*? In order to answer this question, we first need a way to determine how a child has categorized a predicate of this sort. One approach would be to put a predicate the child has just learned into a sentence with an expletive subject (cf. (9-10)) and ask if the child finds the sentence acceptable.

(11) It gorps to be raining.
(12) It is daxy to please John.

If sentences (11-12) are acceptable, then the predicates *gorp* and *daxy* do not select an Agent or Experiencer subject. On the other hand, if these sentences are unacceptable, then these predicates require a semantic subject and do not tolerate an expletive.

As noted in section 1.1, some researchers have successfully elicited metalinguistic grammaticality judgments from children (McDaniel &

Cairns, 1996). However, I have been unsuccessful in eliciting such judgments about sentences like (11-12) from children as old as 8 or 9 years. My hypothesis is that with these types of syntactic constructions, children have difficulty with the metalinguistic aspect of the task, and that if we could assess their *implicit* judgment of acceptability we would be able to learn about how they categorize these abstract predicates.

2. Methodology

The question raised immediately is how to assess children's implicit knowledge. To do this, I will tap into something that developmental scientists have known about and exploited for many decades: babies exhibit surprise when they are faced with something unexpected. Exploiting this surprise response, researchers have been able to learn about babies' categorical perception as well as other aspects of cognitive development, by measuring their sucking rate or heartrate (Eimas et al., 1971). In addition, Spelke and her colleagues (Spelke, 1988, 1991; Spelke et al., 1995; Woodward et al., 1998) measured relative looking time as an indication of how infants expect objects to move through space. Babies look longer at a scene in which inanimate objects move by themselves compared to one in which the objects are pushed by something else, or in which people move by themselves. Using the same principle, Onishi & Baillargeon (2005) assessed 15-month-olds' expectations about where an observer should look for a hidden object according to the knowledge the observer had about the object's location. Babies looked longer at the scene in which a person searched for an object in a location incongruent with where they could have known the object to be.

Measuring relative looking time has also been used in a variety of linguistic studies, in which experimenters measure how long children turn toward one sound source or another, as in the Head-Turn Preference Procedure (HPP; Werker et al., 1981; Werker & Tees, 1984; Jucszyk et al., 1992), or how long they look at one scene or another after hearing a linguistic prompt, as in the Intermodal Preferential Looking Paradigm (IPLP; Hirsh-Pasek & Golinkoff (1981, 1996). These methodologies have been used successfully in hundreds of studies and reveal important facts about how babies perceive and interpret language. But they cannot be used to tell us how 3- and 4-year-olds judge the acceptability of sentences. The HPP procedure cannot be used for this because it was developed specifically to be used with very young infants; the maximum age with which it has been reported to be used is 18 months (Santelmann et al.,

2003).[2] Moreover, it measures babies' perception of a difference between two stimuli, and a pair of sentences that differ in their grammaticality will necessarily be different, purely in terms of their surface form. Recognizing a difference between them does not tell us whether the child finds their *grammatical representation* different. Similarly, the IPLP cannot be used to assess a child's judgment of the form of a sentence because, by its nature it does not measure a child's judgment about a linguistic form, as such. Rather, it tells us how a child interprets the semantics of the sentence — whether the sentence matches or does not match a scene. Thus, neither of these methodologies is suited to the problem at hand.

2.1 Reaction Time

Reaction Time has been widely used as a means of measuring how adult speakers process language (Rubenstein et al., 1970; Meyer & Schvaneveldt, 1971; Meyer et al., 1975; Bates & MacWhinney, 1993; McElree & Griffith, 1995, i.a.) and other aspects of cognition (Luce, 1986). It is related to the familiar "double-take" that occurs when we encounter an ungrammatical sentence or a garden path. This double-take often translates into a longer time to respond to a prompt, or to process the input. Capitalizing on both the fact that children indicate surprise when presented with something unexpected, and that ungrammatical input is unexpected and generally yields a slower response, my hypothesis is that we can assess whether a given sentence/question is grammatical or ungrammatical for children by measuring how long it takes them to respond to the prompt.

I noted in the introduction that this methodology had not previously been used to assess children's grammatical knowledge. In fact, RT has been used in a handful of studies on language acquisition, both on L1 (Corrigan, 1988; Naigles et al., 1995) and L2 (Bley-Vroman & Masterson, 1989). But these studies measured subjects' time to perform a metalinguistic task, rather than to assess their degree of surprise. My objective is to eliminate the metalinguistic aspect of the grammaticality judgment task. Therefore, my proposal is to pose a series of yes/no questions, some grammatical and some ungrammatical, and measure how long it takes children to answer each question. The assumption is that, on average, ungrammatical questions should be answered more slowly compared to grammatical questions.

[2] According to M. Soderstrom (personal communication), HPP can be used with babies up to 30 months of age, but there are no published studies with children of this age.

2.2 Procedure

First, I will describe the general procedure, including the recording and importing of data, and the analysis. Then I will discuss how the methodology works specifically in the case of novel abstract predicates.

Children were videotaped with their faces fully visible. An external microphone recorded the sound. After watching a short video of toys or puppets interacting, children were asked two questions by a puppet. One of the questions should have sounded grammatical, and the other should have sounded ungrammatical due to an argument structure violation, as in the example in (13) (order of presentation was counterbalanced across items and across participants).

(13) A:Hey! The policeman is sleeping!
B:Really? The policeman is sleeping?
A:Yeah! The policeman is sleeping.
B:Wow! Is the nurse sleeping too?
A:No, the nurse is not sleeping.

Question 1: Is the nurse sleeping? (grammatical; target answer "no")
Question 2: Is the policeman sleeping the nurse? (ungrammatical)

I recorded children's responses and, although the dependent variable of interest is RT, not correctness, I also assessed children's correctness for the grammatical questions.

RT can be measured using a stopwatch, which is how it was measured in the study by Naigles, Fowler & Helm (1995). However, the stopwatch method is subject to human error, both in starting and stopping the clock. Even without making an error, humans exhibit a small lag in actually executing the start/stop action, which can add as many as a couple of hundred milliseconds to the recorded times. A more precise way to measure RT involves analyzing the video footage using the ELAN program. ELAN is software that can be used for annotating and analyzing audio and video data. It is freely available through the Max Planck Institute (http://tla.mpi.nl/tools/tla-tools/elan/). In order to analyze the data with ELAN, we created an iMovie file for each child and imported the question and answer session for each item into iMovie. We then exported this "movie" using Quicktime to create a .mov file, and by changing the export setting to "sound to wave" we created a .wav file for the audio portion.

With the audio (.wav) and video (.mov) files in place, after opening ELAN one can open a new file for each child and then add the media files

(.wav and .mov) as well as a template for annotating the recording. The template is not necessary but is helpful in coding the responses according to different types of stimuli. For our template we used the following items: a code for the kind of stimulus (warm-up vs. filler vs. target), the kind of predicate (for target items), whether the question was grammatical or ungrammatical, and whether the child answered *yes* or *no*.

Once the ELAN project is opened, the coder can view the video of the session and the soundwave of the audio file, as well as hear the sound simultaneously while watching the video. The coder can then select (highlight) the interval from the end of the experimenter's question until the beginning of the child's answer, and annotate that selection with items from the template. The waveform can be helpful in identifying with precision the end of the experimenter's speech and the beginning of the child's response. Researchers must decide in advance how to deal with gestural (nonverbal) responses. We chose to include these responses, coding them such that the beginning of the child's response was the point where the child's head movement was an unambiguous nod (for yes) or shake (for no).

Figure 1 shows a screenshot of what a portion of the video and audio data looks like in ELAN.

Once the entire file has been coded, the annotations can then be exported as tab-delimited text so that they can be viewed in Excel or another spreadsheet program, in which one can see the start time, end time and duration of the interval (in milliseconds), along with all of the information that was coded using the template. The exported data can be combined into a single spreadsheet file and, if each file is labeled with an identifier for the participant, the participant ID will be associated with each line of data. Durations can be easily converted to log10 transformed durations, as is standard in analyzing RT data.

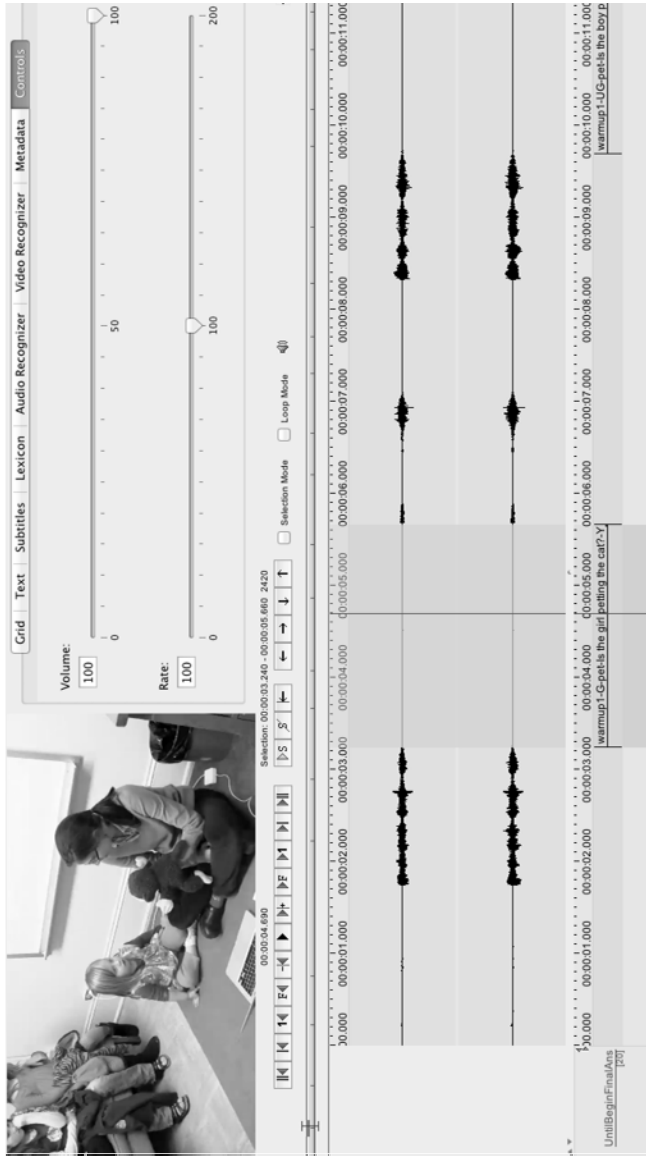

Figure 1. Screenshot of selected interval in ELAN.

2.3 Results

I have conducted two experiments using this methodology, one with children aged 4 to 7 years and the second with children aged 3 to 4 years (both are described in Becker, 2015a). First, I present the results of the warm-up and filler items to show that the basic premise of the methodology holds: children generally take longer to answer ungrammatical questions than grammatical ones. Figure 2 shows 4 to 7-year-old children's RT in answering grammatical and ungrammatical questions involving the familiar verbs *play* and *borrow* (e.g. *Did the farmer play with the car?* vs. **Did the farmer play the car to his friend?*) and two novel verbs, one used intransitively (*ballop*) and the other transitively (*zorp*).

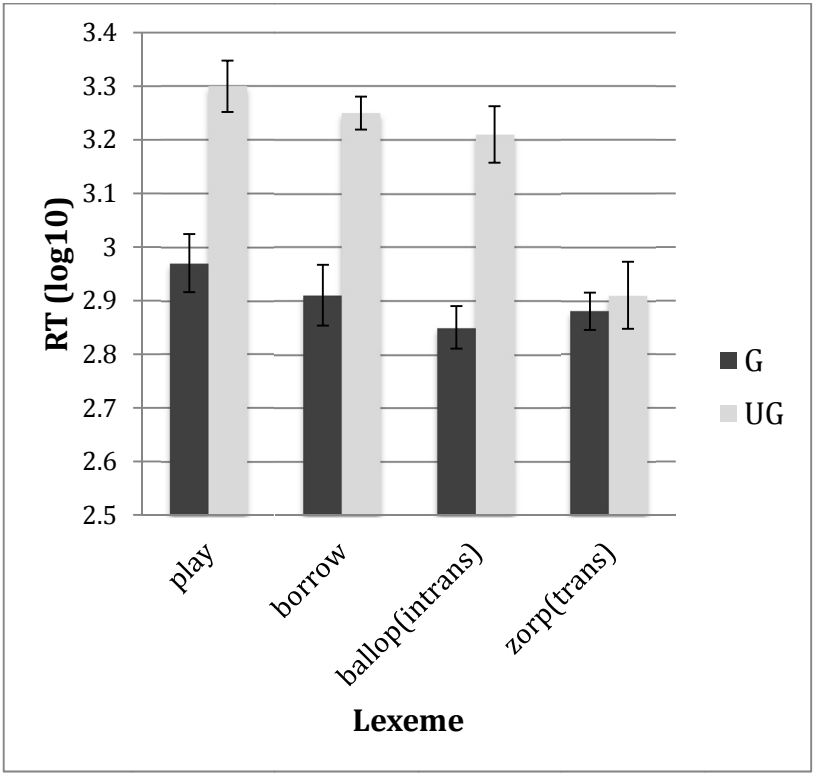

Figure 2. Experiment 1: children aged 4-7 years. G = Grammatical, UG = Ungrammatical