

THE ROLE OF SUPRALEXICAL PROSODIC UNITS IN SPEECH PRODUCTION:
EVIDENCE FROM THE DISTRIBUTION OF SPEECH ERRORS

by

WOOK KYUNG CHOE

A DISSERTATION

Presented to the Department of Linguistics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2013

DISSERTATION APPROVAL PAGE

Student: Wook Kyung Choe

Title: The Role of Supralexical Prosodic Units in Speech Production: Evidence from the Distribution of Speech Errors

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

Melissa A. Redford	Chairperson
Eric Pederson	Core Member
Vsevolod Kapatsinski	Core Member
Kaori Idemaru	Institutional Representative

and

Kimberly Andrews Espy	Vice President for Research and Innovation; Dean of the Graduate School
-----------------------	--

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded December 2013

© 2013 Wook Kyung Choe

DISSERTATION ABSTRACT

Wook Kyung Choe

Doctor of Philosophy

Department of Linguistics

December 2013

Title: The Role of Supralexical Prosodic Units in Speech Production: Evidence from the Distribution of Speech Errors

The current dissertation represents one of the first systematic studies of the distribution of speech errors within supralexical prosodic units. Four experiments were conducted to gain insight into the specific role of these units in speech planning and production. The first experiment focused on errors in adult English. These were found to be systematically distributed within the highest-level supralexical prosodic unit, the Intonational Phrase (IP), providing evidence for its psychological reality. The specific distribution of errors—fewest in unit-initial position, with a gradual increase in errors across the unit—was interpreted to suggest that the IP functions as a planning domain: the unit is activated as a whole, and activation gradually decays with time leading to an increase in errors. The second experiment was motivated by the idea that a decrease in IP activation is best understood in the context of working memory processes. Children's speech was examined in preference to adult speech because it is less automatized and so likely more influenced by working memory. The findings were that children with better working memories produced shorter IPs and relatively more anticipatory errors than children with poorer working memories. The results provided further evidence for the role of IPs in planning. The third and fourth experiments extended the investigation to

another language, Korean, and examined the role of a mid-level prosodic unit, the Accentual Phrase (AP), in planning and production. The results indicated the same pattern of error distribution in the Korean IP as in the English IP. In contrast, more errors occurred in AP-initial position than in the second half of the unit, and the elicited errors tended to preserve AP-internal structure. The results were interpreted to suggest that the AP provides a structural frame within which elements are slotted for production. Overall, the results are consistent with the idea that these units play a critical role in the planning and production process. The results also suggest that different units within the prosodic hierarchy function differently: the IP functions as a planning domain, and mid-level units (i.e., AP) provide the structure needed to accomplish serial ordering in speech.

This dissertation includes previously published co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Wook Kyung Choe

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
Ewha Womans University, Seoul, South Korea

DEGREES AWARDED:

Doctor of Philosophy, Linguistics, 2013, University of Oregon
Master of Arts, English Language and Literature, 2006, Ewha Womans
University
Bachelor of Science, Biological Science, 2004, Ewha Womans University

AREAS OF SPECIAL INTEREST:

Psycholinguistics: Speech production and planning
Experimental phonology and phonetics
Language acquisition

PROFESSIONAL EXPERIENCE:

Research Assistant to Professor Melissa A. Redford, Speech and Language Lab,
University of Oregon, 2009-2013

Teaching Assistant, University of Oregon, 2008, 2011

Research Assistant to Dr. Linda Forrest, Center for Applied Second Language
Studies (CASLS), University of Oregon, 2008-2009

Research Assistant, Brain Korea for Department of English Language and
Literature, Ewha Womans University, 2004-2006

GRANTS, AWARDS, AND HONORS:

Graduate Research Fellowship, University of Oregon, 2008-2013

Graduate Teaching Fellowship, University of Oregon, 2008, 2011

Research Assistant Fellowship, Ewha Womans University, 2004-2006

Academic Scholarship, Ewha Womans University, 2002

PUBLICATIONS:

Choe, W.K. & Redford, M.A. (2013). The effect of interpretation bias on the production of disambiguating prosody. *Proceedings of the 21st International Congress on Acoustics, (Montreal, Canada)*. International Commission for Acoustics.

Choe, W.K. & Redford, M.A. (2012). The distribution of speech errors in multi-word prosodic units, *Laboratory Phonology, 3(1)*, 5-26.

Choe, W.K. (2011). The distribution of speech errors in prosodic phrases in Korean. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Science (ICPhS-11 Hong Kong)* (pp. 480-483).

ACKNOWLEDGMENTS

Although I believe that this dissertation is a milestone in my academic career, it is difficult to imagine finishing the doctoral course without a tremendous amount of support and encouragement from many people on both sides of the Pacific Ocean.

First of all, I would like to express my sincere gratitude to my advisor, Professor Melissa A. Redford for her indescribable support and advice. Her profound insight and enthusiasm as a researcher have always inspired me. With her invaluable guidance and suggestions, I learned how to research, write, give a talk, discuss ideas and much more. She continually challenged me to understand the big picture as well as the details of my research. The time we spent together made me an academically and mentally strong researcher. She has supported me in many areas—emotionally, financially (Award number R01HD061458 from the NICHD) and, needless to say, intellectually. She was, is and will always be my teacher, mentor, family and good friend.

My deepest thanks go to the members of my committee: Professor Eric Pederson for his thorough comments and suggestions on the empirical approach of this dissertation, Professor Vsevolod Kapatsinski for his perceptive insight to help me better understand and analyze the data, and Professor Kaori Idemaru for her continuous warm encouragement and valuable comments. Their precious time and critical feedback have greatly improved this dissertation.

My sincere appreciation is extended to other faculty members in the Linguistics program at the University of Oregon. They have provided exceptional guidance and encouragement during the past six years. I also want to thank Korean linguists for their support and guidance. Professor Eun Jin Oh, my Master Thesis advisor and mentor,

deserves special thanks for all her advice and help. She has showed me the right direction as a researcher and encouraged me to move forward.

My hearty appreciation goes to my colleagues and friends in this program: Brian, Charlie, Hema, Hideko, Irina, Jason, Julia, Paul, Shahar, and Ying. Thanks to their warm encouragement, intellectual knowledge and good care of me, I have never felt alone throughout my six years in Eugene. My special thanks go to Grace and Prakaiwan for being my stress-relief channels. The two ladies are one of the most precious gifts that I received from my graduate school experience. I would also like to express my gratitude my lifelong friends in Korea for their warm encouragement. All their support and friendship from a distance has cheered me up to complete this journey. My thanks also go to all the Korean catholic church members in Eugene for their prayers and warm concern.

During this long and challenging journey, I would have been lost without the emotional support from my family. Words alone cannot adequately express how grateful I am for the unconditional love, trust, and support that my parents have given. It could not have been easy to encourage me from so far away, but they did so with patience and grace. I also thank my sister for being with me during my lonely life in Eugene. I want to extend my heartfelt thanks to my relatives and friends in California and in Korea. They have always welcomed me and provided generous support.

Last but not least, I would most especially like to thank my husband, colleague and friend, Dr. Kyoung Ho Kang. He was the first reader of all the drafts of this dissertation and provided valuable comments. He patiently held my hand through all the ups and downs along this journey and constantly expressed admiration for what I have done. It would have been impossible to complete this dissertation without his support and love.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
II. BACKGROUND.....	5
2.1. Speech Errors as Data	5
2.2. Speech Planning and Production Models	9
2.2.1. Scan-and-Copy Model	11
2.2.2. Spreading Activation Model.....	13
2.2.3. The Extended Spreading Activation Model.....	15
2.2.4. Summary and Implications for the Role of Supralexical Prosodic Units in Production	18
2.3. Distributional Patterns of Speech Errors.....	20
2.4. Prosodic Phrases as Speech Processing Domains.....	26
2.5. Current Dissertation.....	30
III. EXPERIMENT 1: SPEECH ERRORS IN REAL-WORD TONGUE TWISTERS PRODUCED BY ADULT ENGLISH SPEAKERS	32
3.1. Introduction.....	32
3.2. Methods.....	36
3.2.1. Participants.....	36
3.2.2. Stimuli.....	36
3.2.3. Procedure	36

Chapter	Page
3.2.4. Coding.....	38
3.3. Results and Discussion	40
3.3.1. Speech Errors	40
3.3.2. Prosodic Boundaries	41
3.3.3. Speech Errors and Prosodic Boundaries.....	43
3.4. General Discussion	49
3.4.1. Big and Small IUs: One Unit or Two?.....	50
3.4.2. The IP as a Speech Planning Domain.....	52
 IV. EXPERIMENT 2: SPEECH ERRORS IN REAL-WORD TONGUE TWISTERS PRODUCED BY SCHOOL-AGED ENGLISH-SPEAKING CHILDREN	54
4.1. Introduction.....	54
4.2. Methods.....	57
4.2.1. Participants	57
4.2.2. Standardized Assessment of Working Memory	58
4.2.3. Elicitation: the Tongue-Twister Task	59
4.2.4. Coding.....	61
4.3. Results and Discussion	62
4.3.1. General Characteristics of Children’s Speech Errors	62
4.3.2. The Effect of Working Memory on Speech Planning	67
4.4. General Discussion	69

Chapter	Page
V. EXPERIMENT 3: SPEECH ERRORS IN REAL-WORD TONGUE TWISTERS PRODUCED BY KOREAN SPEAKERS	73
5.1. Introduction.....	73
5.2. Methods.....	76
5.2.1. Participants.....	76
5.2.2. Stimuli.....	76
5.2.3. Procedure	77
5.2.4. Coding.....	77
5.3. Results and Discussion	78
5.3.1. Speech Errors	78
5.3.2. Supralexical Prosodic Units.....	79
5.3.3. Distribution of Errors in Supralexical Prosodic Units	81
5.4. General Discussion	86
5.4.1. Overall Patterns of Speech Errors in Korean.....	87
5.4.2. Korean Supralexical Prosodic Units in Speech Planning and Production.....	89
VI. EXPERIMENT 4: SPEECH ERRORS IN NONSENSE TONGUE TWISTERS PRODUCED BY KOREAN SPEAKERS	92
6.1. Introduction.....	92
6.2. Methods.....	96
6.2.1. Participants.....	96
6.2.2. Stimuli.....	96

Chapter	Page
6.2.3. Procedure	99
6.2.4. Coding.....	100
6.3. Results and Discussion	101
6.3.1. General Characteristics of Speech Errors	101
6.3.2. Effect of Error-Prone Syllable Position on Error Occurrence	101
6.3.3. The Distribution of Target Syllable Errors within the IP	103
6.4. Conclusions.....	105
VII. CONCLUSIONS AND FUTURE DIRECTIONS	107
7.1. Distribution of Speech Errors in Supralexical Prosodic Units.....	107
7.2. Planning Domains versus Planning Frames.....	109
7.3. Future Directions	113
APPENDICES	117
A. ENGLISH TONGUE TWISTERS AND MATCHED NONTONGUE TWISTERS	117
B. KOREAN TONGUE TWISTERS	119
REFERENCES CITED.....	121

LIST OF FIGURES

Figure	Page
2.1. Phonological encoding in the production of the sentence in Dell's (1986) spreading activation model	14
2.2. Dell, Burger, and Svec's (1997) spreading activation model of language production	16
3.1. A schematic of possible patterns of error distribution within supralexicall prosodic units	33
3.2. Position of words containing speech errors as a function of the number of words in utterances, big IUs, and small IUs	44
3.3. Position of words containing speech errors as a function of the number of words in sentence-nonfinal and sentence-final IUs	45
3.4. Number of errors in medium and long IUs as a function of the position of the words containing speech errors.....	47
4.1. Position of words containing speech errors as a function of the number of words in the IUs	65
4.2. Number of errors in medium and long IUs as a function of the position of the words containing speech errors.....	66
4.3. Average length in words of error-containing IUs as a function of the standardized NR scores.....	68
4.4. AProp values as a function of the standardized NR scores	69
5.1. Position of syllables containing speech errors as a function of the number of syllables in the IP, ip, and AP	82
5.2. Number of errors as a function of error position in the AP and position within the ip/IP	83
5.3. Number of anticipatory and/or perseveratory errors in ip/IP as a function of error position.....	85
6.1. Two possible roles for the AP in speech planning.....	94

6.2. The mean number of target syllable errors as a function of syllable position 102

6.3. The mean number of target syllable errors for each AP position within the IP.... 104

LIST OF TABLES

Table	Page
3.1. Sentences with different types of internal prosodic boundaries as a function of sentence length and whether the sentence was produced with or without speech errors	42
4.1. Mean number and length in words of big and small IUs in a single repetition of a tongue twister as a function of contextual-error occurrence.....	64
5.1. Number and length of the 3 different supralexicalexical prosodic units that contained one or more contextual errors.....	80
6.1. Four types of tongue twisters and the precise manipulation of three factors.....	97

CHAPTER I

INTRODUCTION

Fluent speech depends in large part on a speaker's ability to plan speech in advance of speaking. To take just one example, the choice of the indefinite article *a* or *an* depends on knowing at some level that the first segment of a following word is a consonant or vowel. Planning processes depend on working memory (Baddeley, 1998). Material is retrieved from long-term stores and remains activated while being readied for speech execution. Given the complexity of language, it should not be surprising that planning occurs at many levels during the production process: from syntactic planning, where syntactic constructions are selected (e.g., Bock & Warren, 1985; Prat-Sala & Branigan, 2000; Solomon & Pearlmutter, 2004; Allum & Wheeldon, 2007; Gallo, Jaeger, & Smyth, 2008) to phonological encoding, where sound sequences are represented in sufficient detail for speech motor planning to occur, which involves specifying articulatory goals and retrieving the relevant motor routines (e.g., Van der Merwe, 1997; Guenther, Hampson, & Johnson, 1998; Perrier, 2012; Perrier & Fuchs, to appear). The last stages in the planning process, from phonological encoding to articulatory specification, likely makes direct use of the phonological loop component of working memory (Baddeley, 2000). Some portion of the overall message is chunked for the elaborated planning that is necessary for execution. The current dissertation investigated the relevance of supralexical prosodic units to these last planning stages. Supralexical prosodic units are defined by intonational contours and delimited by specific tonal and

temporal patterns. They are usually larger than lexical words and often out of alignment with syntactic clauses (Shattuck-Hufnagel & Turk, 1996).

Speech errors provide perhaps the most important source of evidence for the claim that speech is planned in advance of speaking. Speech errors, often called *slips of the tongue*, are unintentional deviation from a targeted linguistic form. These errors have been used to argue both for the units involved in speech planning and for the size of the domain within which planning occurs. For example, Garrett (1980a) noted that word exchange errors, such as *this spring has a seat in it* for *this seat has a spring in it*, involves the same type of grammatical unit in different clausal positions, and that sound exchange errors, such as *a disorder of speech, spictly streaking* for *strictly speaking*, involves units in the same syllable position in words that are adjacent to one another. Garrett took these different error patterns to indicate that syntactic planning and the serial ordering of words occur within the scope of a clause while phonological planning and the serial ordering of phonemes occur within the scope of a syntactic phrase.

Garrett's (1980a) also argued that sound errors demonstrate how the right (or wrong) phonemes are selected and arranged in the right (or wrong) order. Phonological encoding refers specifically to this selection and ordering process, and researchers have continued to rely on specific error patterns to inform theorizing regarding the mechanism by which encoding happens and the role of different linguistic units involved in these processes (e.g., Roelofs, 1999; Goldrick & Blumstein, 2006; Schnur, Costa, & Caramazza, 2006). For example, Shattuck-Hufnagel (1979, 1983) has suggested that sound errors provide evidence that syllable structure provides a frame within which phonemes are ordered (i.e., syllables and phonemes are psychologically real), and that

candidate phonemes are appropriately ordered by virtue of being tagged with information about syllable position (i.e., onsets, nuclei, and codas).

Once sequenced, phonological material is then executed. Theories about how this occurs have also been inspired by speech errors, and specifically by the distribution of errors within units. For example, Dell and his colleagues (1993) have suggested that most errors occur at the beginning of words because the serial execution of phonemes is aided in part by activation from preceding phonemes, and word-initial phonemes do not have a preceding context from which to receive activation.

Most current models of speech production have focused on phonological encoding at the level of lexical words (for reviews see Levelt, 1999; Meyer & Belke, 2007). A few consider encoding up to the level of the phonological word (i.e., one content word with one unstressed function word such as *the boy*) (Shattuck-Hufnagel, 1992; Levelt, Roelofs, & Meyer, 1999). But we know of only one model (Keating & Shattuck-Hufnagel, 2002) that incorporates larger supralexical prosodic units. This is in spite of the phonetic evidence, from articulatory strengthening to pause duration (Fougeron & Keating, 1997; Krivokapić, 2007, 2012; for a review see Shattuck-Hufnagel, to appear), that strongly suggests that speakers generate planning structures that are larger than words, but not necessarily in alignment with clauses.

The focus on word-level production may be due to the fact that studies on naturalistic and experimentally elicited speech errors rarely consider or manipulate units larger than lexical words. This in turn means that only very minimal descriptions exist for how errors are distributed with respect to supralexical prosodic units. Thus, a major practical goal of this dissertation was to provide such data. The theoretical goal was to

obtain insight into what roles supralexicical prosodic units play in speech planning and production. In support of this latter goal, the next chapter presents the background and assumptions on which our theoretical interpretations are based. The data are then presented in the following chapters. Note that the data presented in Chapter III were previously published in a co-authored journal article, and that a portion of the data presented in Chapter V was published in a peer-reviewed proceedings. Overall, the data indicate not only the psychological reality of supralexicical prosodic units such as the intonational phrase and the accentual phrase, but also their different linguistic statuses and roles in planning and production.

CHAPTER II

BACKGROUND

2.1. Speech errors as data

On average, a fluent adult speaker makes one or two speech errors per 2000 words (Levelt, 1999; Roelofs, 1997), but error rates are affected by the redundancy of linguistic materials and various non-linguistic factors including the speaker's age. Psycholinguists have collected speech errors from spontaneous speech from different populations or have elicited errors in experimental settings using repetitive linguistic materials in order to investigate the psychological reality of linguistic units and the speech planning and production processes (e.g., for spontaneously produced errors from adults and children: Fromkin, 1971; Shattuck-Hufnagel, 1979; Garrett, 1976; Jaeger, 1992; Wijnen, 1992; Stemberger, 1989; e.g., for experimentally elicited errors from adults and children: Barrs, Motley, & Mackay, 1975; Shattuck-Hufnagel, 1992; Dell, Burger, & Svec, 1997; Frisch, 2000; Vousden & Maylor, 2006; Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; Stemberger, 2009; Croot, Au, & Harper, 2010). This dissertation experimentally elicits speech errors in order to investigate the role of supralexical prosodic units in these same processes.

Speech errors can be categorized as either contextual errors or non-contextual errors. When the source for an error is identifiable within the same intended utterance, the error is a contextual error (Stemberger, 1985; Dell, 1986; Roelofs, 1996). The source of the error can also come from outside the utterance. These are non-contextual errors. Substitutions (e.g., *oh, you did the wash for you did the dishes*, from Stemberger, 1985),

additions (e.g., *are those are for the taking* for *are those for the taking*, from Fay, 1982), and deletions (e.g., *retch your legs* for *stretch your legs*, from Fromkin, 1973) are all examples of non-contextual errors¹. Insofar as our goal is to investigate how speakers plan speech, and because it is difficult to assess the proximal cause of non-contextual errors, only contextual errors will be considered in this dissertation.

Contextual errors are also referred to as interaction errors (Shattuck-Hufnagel, 1992), movement errors (Vousden, Brown, & Herley, 2000) or serial-order errors (Dell, et al., 1997). There are three types of contextual errors: anticipatory, when the source is subsequent to the error; perseveratory, when the source precedes the error; exchange, when two elements exchange their position within the same phrase or utterance².

Examples in (2.1) below, all from Fromkin (1971), illustrate these three types of contextual errors.

¹ Some studies have proposed that substitutions, additions, and deletions are error types, while anticipations and perseverations refer simply to the direction of errors (Shattuck-Hufnagel, 1979; Stemberger, 1985). However, more often, the two are considered different classes of errors (contextual versus non-contextual), especially in studies concerned with phonological encoding where identifying the source of an error is important to the theoretical goals of understanding the units involved in planning and the serial ordering process (Dell, 1986, Dell et al., 1997; Shattuck-Hufnagel, 1992; Stemberger, 2009). As noted, the current dissertation will focus only on contextual errors for these reasons, but we will treat addition and deletions as contextual errors when it is possible to find the error source within an utterance

² Previous research on speech errors indicates that categorizing contextual errors according to the location of the source is not always easy (Stemberger, 1989; Dell et al., 1997; Pfau, 2009). For example, Stemberger (1989) noted that the sources for some speech errors in his corpus were unclear, especially when possible sources were located before and after the error (e.g., *She shell sea-shell* instead of *sells*). When the source was not easily identified as preceding or following the error, Stemberger classified the error as both anticipatory and perseveratory. The specific criteria of distinguishing different types of errors for this dissertation will be stated in the methods section of each experiment.

- (2.1) a. Anticipatory: week long race → reek long race
b. Perseveratory: John gave the boy → John gave the goy
c. Exchange: keep a tape → teep a cape

Since errors in speech production are systematic rather than random (Fromkin, 1971; Dell, 1986), their occurrence provides insights into how speech is planned and produced in several ways. First, anticipatory errors clearly demonstrate that speakers look ahead and plan their speech prior to actual articulation. These errors are understood to arise when information from upcoming units interferes with the execution of current units. Take, for example, the error *tame time* for *same time* (from Shattuck-Hufnagel, 1987). The clear source for the error *tame* is *time*, providing evidence that the speaker preplanned the noun phrase, and that the onset segment /t/ in *time* was activated and articulated earlier than it was supposed to be.

Second, systematic error patterns provide evidence for linguistic units involved in the speech planning and production processes. Fromkin (1971) was one of the first to use the structure preserving nature of errors as evidence for the psychological reality of units ranging from phonetic features to syntactic phrases. Take, for example, the spontaneously produced speech errors such as *phi-so-lo-phy* for *phi-lo-so-phy*, and *furger surgery* for *further surgery* that she describes. These types of error suggest that syllables and segments are units of speech planning and production. For example, in *phi-so-lo-phy* the error involves the exchange of whole syllables: in *furger surgery* for *further surgery* there is the early articulation in syllable onset position within a word of a word-medial syllable onset from the subsequent word.

Finally, speech error data have provided us with evidence that different levels of linguistic representation are relevant to planning. For example, Garrett (1976, 1980a, 1980b) noted that word exchange errors, such as *I left the briefcase in my cigar for the cigar in my briefcase*, preserves the grammatical categories and functions of their sources even while they occur in different syntactic phrases within the same clause. This suggests planning at the syntactic level. On the other hand, he found that sound exchange errors, such as *even the best team lost/s/ for team/z/ lost*, frequently take their sources from adjacent words rather than from the words in the same grammatical category, and they occur in the same syntactic phrase. These types of errors suggest planning at a phonological or phonetic level. Garrett characterized syntactic planning as planning at the ‘functional level,’ and noted that the relevant planning domain was the clause. He characterized phonological or phonetic planning as planning at the ‘positional level,’ and hypothesized that the relevant planning domain was the syntactic phrase.

In sum, specific error characteristics or patterns provide information for how speech is planned and produced and what units are involved in the planning and production processes. The errors have therefore provided the foundation for a number of speech planning and production models. In the next section, we review three of these models and consider the specific insights that they have brought to our understanding of phonological encoding in advance of speech production.

2.2. Speech planning and production models

In the context of spoken language production, the terms ‘speech production’ and ‘language production’ are often used interchangeably³. This is perhaps because it is difficult to decide at what point along the continuum of spoken language production ‘language’ shifts to or separates from ‘speech’. That said, models of ‘language production’ (e.g., Butterworth, 1980; Stemberger, 1985, 1991; Levelt, 1989) generally explain the spoken language production process as starting from the level of conceptual-semantic representation, where non-linguistic meanings or preverbal messages are generated and then encoded into linguistic forms. By contrast, models of ‘speech production’ often begin with the assumption of a morpho-syntactic plan and then consider how lexical selections, morpho-phonological formulations, articulatory planning, and implementation take place (for a review, see Meyer & Belke, 2007).

Despite the tendency for ‘language’ and ‘speech’ production models to focus on different stages in the production process, researchers have also used the two terms interchangeably and without clear definition. The one exception to this generalization that I know of is Fowler (2007), who differentiates theories of language production from theories of speech production in a footnote, claiming that the former refers to “planning for and implementation of meaningful utterance,” whereas the latter does “planning for and implementation of language forms (Fowler, 2007, p. 489).” The term ‘speech production’ will be used here to refer to the process of planning, serially ordering, and

³ In two review papers about the production of spoken language (Levelt, 1999; Meyer & Belke, 2007), the authors cite a total of 61 articles or book chapters whose titles include either the phrase ‘language production’ ($N = 24$) or ‘speech production’ ($N = 37$). Regardless of the title, most of the articles considered production processes from the level of grammatical encoding onwards.

executing phonological representations that have already been activated in the lexicon⁴.

These final stages of production, after syntactic frames have been built and lexical retrieval has occurred, represent the scope of the current dissertation.

Although supralexicalexical prosodic units are the units of interest in the current dissertation, the largest linguistic unit that most existing models of speech production assume is either the prosodic or lexical word. This section provides a detailed review of 3 such models and discusses the similarities and differences between them. The goal of the review is to understand how different models account for the selection of the right phonemes in the right order. We will then build on specific ideas put forth in these models to make predictions about the distribution of contextual errors in supralexicalexical prosodic units.

Before moving onto the details of the individual models, note that those reviewed here assume that the phoneme is the fundamental unit of phonological encoding. This assumption is supported by the frequency with which phonemes participate in speech errors. Spontaneous speech error data show that errors involving a single segment (e.g., *some kunny kind* instead of *funny*) are strikingly more frequent than errors involving a single phonological feature (e.g., *glear plue sky* instead of *clear blue*). Additional evidence for the phoneme as a fundamental unit comes from “implicit priming” experiments (Roelofs, 1999), which show that a prime that includes the same phoneme as a response word leads to faster reaction times than when the prime merely shares

⁴ Some theories of speech production focus on a stage that may be further along in the planning-production process; namely, the selection and execution of gestures, which define the articulators brought to bear to achieve specific vocal tract constrictions (e.g., Byrd & Saltzman, 2003; Fowler, 2007). These models have been built to explain articulatory timing patterns, which are well-defined in the kinematic and acoustic realms. With the exception of recent work by Goldstein et al. (2007), work on articulatory timing has rarely concerned itself with explaining speech error data.

phonological features or a whole syllable with the response. Adopting the assumption of the phoneme as a fundamental unit is also convenient because it allows for conclusions regarding planning based on transcribed/perceptual data. Finer-grained acoustic/kinematic data must be used to evaluate the importance of features or articulatory gestures to planning and production. It is this convenience factor in particular that leads me to also adopt the assumption that phonemes are the fundamental unit of phonological encoding.

2.2.1. Scan-and-copy model

Shattuck-Hufnagel (1979, 1983) proposed a scan-and-copy model to account for the structure preserving nature of speech errors⁵. In brief, the model assumes that phonemes are retrieved, held in a buffer (working memory), and then placed in pre-generated slots that are specified by the prosodic structure. She also proposed that phonemes are independent of the slots they fill. The assumption of segment-structure independence explains the structure preserving nature of contextual errors. For example, the error in the sentence *change the first part* (instead of *first part*) is only explained if a target slot in the word *first* (i.e., an initial slot) can be assumed to have been replaced by a phoneme designated for a later slot. If the target and slot were to move together, then no slot would be available for the correct rendition of the later word (e.g., the error would be *pirst art* instead of *pirst part* for *first part*). Regarding the notion of frames, Shattuck-Hufnagel is clear that syllable structure is important based on her finding that only 1.9% of exchange errors in the MIT-CU corpus of spontaneous speech errors (Shattuck, 1975)

⁵ The scan-and-copy model is also called a “slot-and-filler” model. These two terms will be used interchangeably to refer to Shattuck-Hufnagel’s model of speech production in the current dissertation.

transposed phonemes from different syllable positions. In the 1983 chapter and in later work (1992), Shattuck-Hufnagel suggests that syllables are themselves organized into larger frames, referencing in particular the metrical foot.

Speech errors arise in the scan-and-copy model due to mistakes in the serial ordering process, which depends on a scan-copier mechanism. First, a candidate set of syllable-structure-tagged phonemes is retrieved from the lexicon and stored in the buffer. Next, a prosodic frame is generated based on already-activated lexical information. The scan-copier is responsible for associating segments with the correct slot in the frame. The scan-copier, working from left to right, selects a target phoneme and copies it into its designated slot. Errors occur when, by chance, the scan-copier mis-selects a segment (e.g., /p/ in *part*), and so associates it with the wrong slot (e.g., the onset slot in the word *first*).

Unlike exchange errors that are solely attributed to the malfunction of the scan-copier, Shattuck-Hufnagel (1983) suggested that anticipatory and perseveratory errors arise due to the malfunction of another critical mechanism, the check-off monitor. After the scan-copier copies a selected segment for each slot, the monitor is meant to delete the segment or mark it as having been used, which prevents it from being re-selected. In the case of anticipatory and perseveratory errors, the monitor fails to delete or mark off the used segment, and so it remains available to the scan-copier as it continues to operate from the leftmost slot of the frame to the rightmost slot.

In sum, Shattuck-Hufnagel's (1979, 1983) scan-and-copy model, as one of the earliest formal models of word-level phonological encoding, established the assumption that contextual errors arise due to mistakes in the serial ordering process, and that specific characteristics of error patterns can tell us how the serial ordering mechanism works. The

model also makes a fundamental assumption regarding planning, namely, that a set of phonemes is retrieved and temporarily stored in a short-term buffer. It is largely due to the concurrent availability of all to-be-executed phonemes that speech errors occur. I will build on the idea of a buffer, which makes explicit the importance of working memory to planning and production, when we consider how supralexicial prosodic units are involved in the planning and production processes.

2.2.2. Spreading activation model

Shattuck-Hufnagel's (1979, 1983) basic proposal that a frame structure underlies phonological encoding was repeated in a number of later psycholinguistic models of speech production. The novel contribution of these later models was to offer alternatives to the scan-copier mechanism and check-off monitor. For example, Dell's (1986) original spreading activation model of speech production assumed independent phonemes and syllable frames, as in Shattuck-Hufnagel (1979), but used spreading activation from higher levels in the plan to lower levels to accomplish ordering during the phonological encoding phase of speech planning. Specifically, in Dell's original model, lexical items are represented as one set of nodes at one level of planning, namely, the morphological level, and these are distinct from phonemes that exist as another set of nodes at a subordinate processing level, namely, the phonological level. These ideas of planning levels and spreading activation will be adopted when we consider aspects of phonological encoding within the domain of supralexicial prosodic units.

Spreading activation is best understood with reference to an example. Let us consider phonological encoding for the sentence, "*dogs like bones,*" illustrated in Figure

2.1. When a currently selected node at the morphological level (e.g., *DOG*) receives enough activation from the syntactic level to pass threshold, it spreads some portion of this activation to the connected nodes, which are phonemes, at the phonological level. The spreading activation model further assumes that phonemes are serially ordered at the phonological level within syllables that are generated by rule (i.e., the structural specification presented in the left column in Figure 2.1). Specifically, at the level of phonological encoding, a syllable frame is generated at the same time that phonemes are activated and tagged for syllable position.

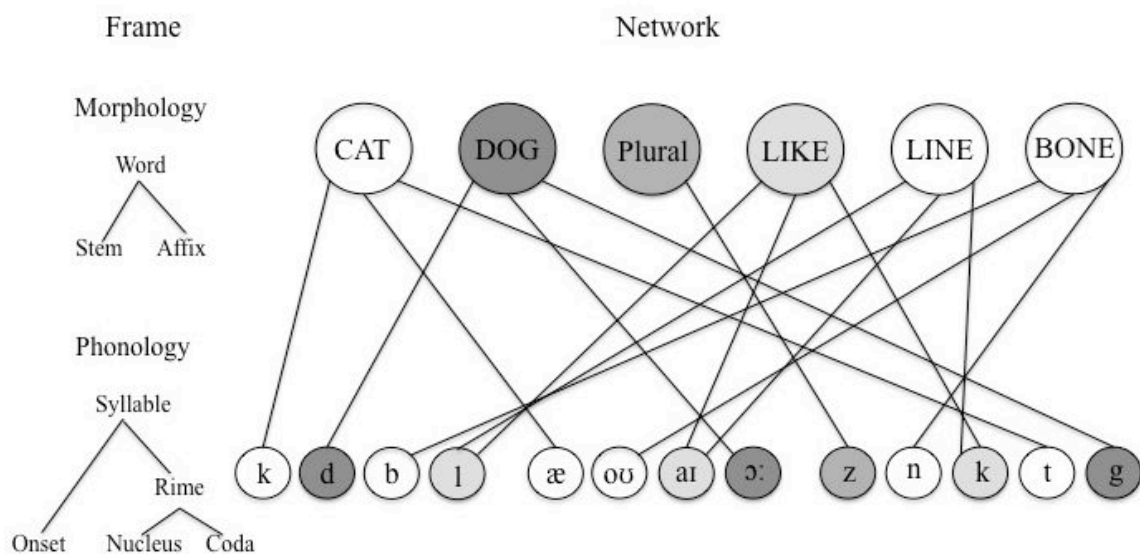


Figure 2.1. Phonological encoding in the production of the sentence *Dogs like bones* in Dell's (1986) spreading activation model. The nodes in the upper layer represent morphemes and the nodes in the lower layer represent phonemes. The degree of darkness of each circle reflects activation level with darker circles being more activated than the lighter ones. The darkest circle represents the currently selected node(s) at each level of representation.

After phoneme selection and spell-out, the activation of the nodes at the phonological and morphological levels falls back to zero. The decay in activation keeps

the currently produced nodes from being reselected and so prevents perseveratory errors. In this way, a single mechanism – activation – is used to replace Shattuck-Hufnagel's (1983) two mechanisms: the scan-copier and check-off monitor.

2.2.3. The extended spreading activation model

Dell's (1986) frame-based spreading activation model was extended further by Dell, Burger, and Svec (1997) to model the distribution of different types of contextual speech errors. In particular, the extended model was designed to account for the relative proportion of anticipatory to perseveratory errors in different corpora. The claim was that these types of errors provide us with specific details about how we “activate the present, deactivate the past, and prepare to activate the future (Dell et al., 1997, p.123)” during speech planning.

In the extended model, Dell et al. (1997) postulate four planning states, each represented by a node in the network: the plan, the past, the present, and the future. The three temporal states (the past, the present, and the future) are excited by connections from the plan. Importantly, the weight (w in Figure 2.2) of the connections from the plan changes based on the “schedule” of each state. Specifically, the weights of activation from the plan to the past, present, and future states are 0, w , and bw , respectively, where w is a long-term positive weight that can be strengthened; and b is a fixed positive value representing anticipatory activation. In addition to gradient connections from the plan, the three temporal states interact with a structural frame that helps to also determine the order of the specific linguistic units according to grammatical rules. As Figure 2.2 shows, the model assumes that all items associated with the past (X in Figure 2.2) receive no

activation (0) from either the plan or the currently selected node in the structural frame, which allows activation in this unit to decay naturally. By contrast, all items associated with the present (Y) receive full activation (1) from both the plan and the selected structural node, and that all those associated with the future receive some activation (b) from the plan and the currently selected structural node. In this way, signals from both the plan and the structural frame serially order the set of linguistic units: (i) the present is activated so that it can be produced, (ii) the past is turned off (i.e., activation is not maintained) so that it does not continue to be produced, and (iii) the future is prepared so that it can be produced next.

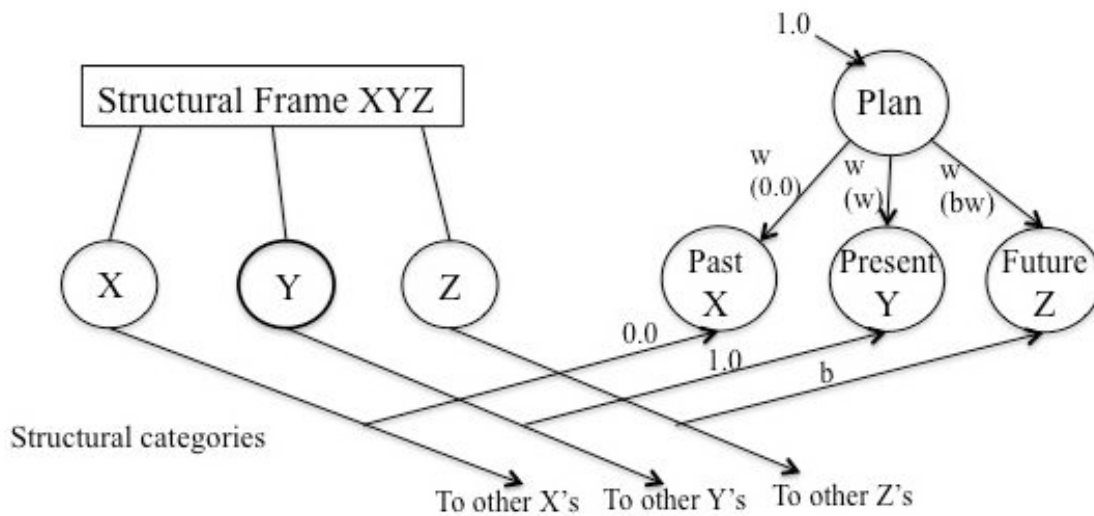


Figure 2.2. Dell, Burger, and Svec’s (1997) spreading activation model of language production that focuses on the serial order problem and accounts for different proportions of anticipatory to perseveratory errors as a function of different speaker attributes (practice, age, pathology). Items (right) for the past, present, and future receive weighted inputs w from the plan and from nodes in the structural frame. The weighted values (in parentheses) are based on the “schedule” associated with each item. The figure indicates that structural node Y is currently selected (adapted from Dell et al., 1997).

In this model, Dell and his colleagues (1997) focused on the effect of practice as one factor that influences the proportion of anticipatory to perseveratory errors that a speaker produces. Specifically, the hypothesis was that practice increases a speaker's ability to prepare the future, but it does not influence the ability to deactivate the past, which is a passive process in any case. When a sequence is practiced, the strength of the connection (w) with the plan increases. Although this strengthening reduces overall error rates by making the present highly active, it also makes the future interfere more with the present because now the future is also more highly activated. This hypothesis was supported by results from elicited speech errors. Participants were asked to produce the same four-word tongue twisters in 8 trial blocks. The results indicated that practice reduced the overall error rates across the 8 trial blocks from 5.9 errors/100 words in the first trial block to 1.5 errors/100 words in the last trial. Practice also changed the proportion of anticipatory to perseveratory errors from 77 anticipatory errors and 135 perseveratory errors in the first trial to 28 anticipatory errors and 23 perseveratory errors in the last trial. The results were taken to suggest that practice did indeed strengthen influences of the "plan" on the serial ordering process.

Dell and colleagues (1997) went on to show that healthy adults also produce lower error rates and a higher proportion of anticipatory errors relative to perseveratory errors than either children or adults who have suffered brain injury. The effect of population differences on both error rates and on the so-called anticipatory error effect was taken to indicate that planning (the strength of the connection between the "plan" node and the temporal states associated with ordering) is limited by intrinsic cognitive abilities such as working memory capacity in addition to being affected by practice. The

importance of intrinsic cognitive abilities to planning at the supralexical prosodic level will also be explored in the current dissertation by investigating errors in children's speech.

2.2.4. Summary and implications for the role of supralexical prosodic units in production

Although all 3 models reviewed here assume that, in order to account for speech errors, we should assume the psychological reality of syllables embedded in some larger frame such as a metrical foot or lexical word, the models differ in their treatment of how phonemes come to be associated with the abstract (empty) frames. Whereas the scan-and-copy model posits a set of phonemes, all equally primed for selection, the spreading activation model assume that some phonemes are more activated than others, and that only phonemes that receive the highest activation are selected during spell out. The different assumptions have consequences for the model architecture. The scan-and-copy model must posit a separate inhibitory mechanism (the check-off monitor) to account for why the same phoneme is not selected over and over again during encoding. By contrast, the spreading activation model relies solely on the relative strength of activation to account for both selection and inhibition. These different architectures also have consequences for the explanation of different kinds of contextual errors. For example, in the scan-and-copy model anticipatory errors result by chance, but in the spreading activation models this result from the influence of different processing levels on phonological encoding. A multi-level architecture is central to the extended model put forth by Dell et al. (1997), which also formalizes the idea that additional information (beyond what is currently being encoded/executed) has been planned. Together,

activation from higher processing levels and the explicit notion of a plan capture the “look-ahead” quality of anticipatory speech errors better than the scan-and-copy model where these types of errors are left to chance. It is also a nice feature of the spreading activation model that both anticipatory and perseveratory errors can be explained with reference to just a single mechanism.

The above review also suggests that it is necessary to incorporate information about the internal structure of the frame to account for the structure preserving nature of speech errors. But the assumption of internal structure may not always be required to account for planning or the distribution of speech errors within the planning domain. For example, control signal models assume that a language-external, clock-like mechanism can be used to represent serial order information (e.g., Houghton, 1990; Burgess & Hitch, 1992, 1996; Vousden, Brown, & Hartley, 2000; but see Hartley & Houghton, 1996 for a model that combines a control signal with frames). No internal structure need be generated, at least within the domain of phonological encoding. Moreover, Vousden and colleagues (2000) argue that the mechanism is sufficient to account for the structure preserving nature of errors, suggesting that their proposed mechanism is powerful. This is encouraging because, as I argue below, the assumption of a fixed internal structure may be problematic when we consider phonological encoding within the domain of supralexical prosodic units.

By definition, supralexical prosodic units are comprised of one or more lexical items and so they vary more in length and phonological make-up than lexical words. This variability is a problem if we assume that an abstract frame requires a fixed internal structure to promote successful phonological encoding. The variability in length and

content suggests that the internal structure of supralexical prosodic units is highly flexible. It also cannot be specified in advance of knowing, for example, the specific lexical stress patterns associated with the items being readied for execution (e.g., see Wagner & Watson, 2010 for a review). It is therefore difficult to imagine how these units could be represented (generated) independently of the encoding process. The assumption that it should be generated independently follows from the assumption that it provides a high-level frame for phonological encoding, and that higher level planning processes hold temporal priority over lower level processes in planning and production (e.g., syntax, then morphology, then phonology). Thus, it may be more reasonable to assume that supralexical prosodic units represent domains of planning rather than planning frames per se, which we understand to describe a unit with internal structure that goes beyond position relative to some boundary.

2.3. Distributional patterns of speech errors

Structure preservation captures the observation that contextual errors appear to preserve the same structural position within a syllable or word as their source. Thus, syllable-onsets are exchanged, anticipated, or perseverated upon, as are offsets and nuclei. However, errors do not occur equally in all these positions. For example, speech errors are most likely to occur in onset position and word-initially than in any other position within the syllable or word (Shattuck-Hufnagel, 1987, 1992; Dell et al., 1993; Vousden, et al., 2000; Stemberger, 2009). Just as structure preservation tells us something about the psychological reality of the linguistic unit, the distribution of errors across positions within a unit suggests something about the specific role that the unit plays in

phonological encoding. The current section reviews unit-specific positional effects on the distribution of speech errors and how the disproportionate occurrence of errors in one position or another has been understood in the psycholinguistic literature.

Vousden et al. (2000) found that 68% of contextual errors in their corpus occurred in word-initial position. The finding that speech errors are more likely to occur in the initial position of a lexical word than in any other position has been referred to as the word-initial effect. Because of the high frequency of monosyllabic words in English, it could be that the word-initial effect is really a syllable-initial effect. However, Shattuck-Hufnagel (1987) differentiated between these two possible effects by investigating patterns in both a spontaneous error corpus and in experimentally elicited errors. She found that 77% of contextual consonant errors in polysyllabic words also involved word-initial consonants. Thus, the word-initial effect does indeed seem to be about the word or lexical item rather than about the syllable.

Shattuck-Hufnagel (1987, 1992) offered one explanation for the word-initial effect that relies on the correlation between initial position and prosodic prominence in English. Specifically, the trochaic pattern in English results in a situation where initial syllables are also typically the head of a metrical foot. Shattuck-Hufnagel (1992) investigated the effect of word-position and prosodic foot structure on error distributions and found that more errors occurred when onsets shared both “word-initial” and “foot-initial” features (e.g., *parrot **fad **food **peril*******) than when they simply shared either “word-initial” or “foot-initial” features (e.g., *parade **fad **foot **parole*******⁶). To account for these findings, Shattuck-Hufnagel postulated an additional, prior stage in phonological

⁶ In these example stimulus word lists, the underlined phonemes are confusable targets and syllables in bold are lexically stressed syllables. All the examples in this paragraph are from Shattuck-Hufnagel (1992).

planning. During this stage, the morpho-syntactic plan provides information relevant to phrasal accents and the stress patterns of lexical items are retrieved and scanned in order to generate a prosodic frame and a set of prosodically-tagged elements that will participate in the serial ordering process:

The organizing frame is made up of prosodic units, each of which contains one or more prosodically prominent syllables (where “prominent” is defined at the phrasal level). In addition, the boundaries of lexical items are marked, left over from the syntactic/morphological structure on which the prosodic structure has been imposed... (p. 245);

and,

...the second component of the planning representation, that is, the short-term store of candidate content-word lexical items, marks these two kinds of information (word boundaries and prominence) in a different way. Since each entry in the table is a single lexical item, the locations of word boundaries are automatically available. In addition, word-onset consonants are marked off as a special subset, since they were not needed during the process of determining prosodic prominence. (p. 246)

The word-initial effect is explained to emerge during a later stage of encoding, when word-onset segments and onset consonants from lexically stressed syllables are separately scanned and selected for insertion into syllable-onset positions in the waiting frame. During this scan, consonants that are tagged for word-onset position and as stress-

associated will appear more similar to one another than those that are either tagged for word-onset position or as stress-associated, which increases the chance of a mis-selection. It is therefore the interaction between word onsets and lexical stress that leads to the word-initial effect.

Dell and his colleagues' take a different approach to explaining the word-initial effect in their Parallel Distributed Processing (PDP) model (Dell, Juliano, & Govindjee, 1993). In this model, structural frames are abandoned and words are instead represented as a sequence of phonemes (similar to the control signal models discussed above). In the PDP model, phonemes are serially executed based on activation from the morphological level and, once execution begins, ordering is preserved by activation that flows from one phoneme to another within a single lexical word. Frequent word-initial speech errors are explained to emerge because these phonemes receive overall less activation than those that also benefit from the activation of a preceding phoneme. If this kind of contextual activation is cumulative across the phoneme string (i.e., the word), then we might also expect that the number of errors within a word would decrease as activation increases with the number of preceding phonemes. Dell et al. (1993) do not report any such gradient effects in the distribution of errors, perhaps because the items that they focus on are monosyllabic words. It could be that gradient error patterns would, in any case, be more obvious in larger units where some kind of cumulative activation would have a greater chance of influencing error distributions. We will keep this possibility in mind when considering serial ordering within the supralexical prosodic domain.

Although most studies that consider error distributions have focused on the lexical word, others have followed Shattuck-Hufnagel's suggestion (1987, 1992) and

investigated the relationship between the distribution of errors and prosodic prominence (e.g., Frisch, 2000). Most relevant to our interest in large prosodic units, Croot, Au, and Harper (2010) investigated the effects of utterance position and prosodic prominence on the distributions of speech errors in four-word utterances. Their findings showed the same word-initial effect described above, but also that errors occurred less often in prominent words (i.e., “emphatically-produced words,” p.433) than in non-prominent words across all position within the utterance. They extended Shattuck-Hufnagel’s (1992) idea to account for this finding, proposing that errors were less likely to occur in the phrase prominent position because only a few (one or two) words per utterance would be tagged in this way.

In addition to finding that prominent items are “protected” from errors, Croot et al. (2010) found that the initial words in their four-word utterances were also less error prone than other words in the utterance regardless of the location of words receiving phrase-level prominence⁷. Croot and her colleagues explained this result by arguing that utterance-initial position has a unique status in speech production. They noted, for example, that this position is also typically a prosodically strong position, and referred to evidence of articulatory strengthening in phrase-initial position to make their case (Fougeron and Keating, 1997; Keating 2006). They further suggested that segments in utterance-initial position likely receive higher activation than those in other positions within the utterance, which reduces the possibility of errors involving these segments.

It should be noted that, by focusing on utterance-initial position, Croot and colleagues (2010) did not explain the other interesting positional effect present in their

⁷ The only exception to this pattern was the prominence condition where the second and the fourth words were accented (w1 W2 w3 W4). In this case, the fewest errors occurred in the second word (see Figure 1 and Table 2 in Croot et al., 2010).

data. Their results also clearly showed that the largest number of errors occurred in the third word of utterances with other errors gradiently distributed across positions. If utterance-initial and other prosodically prominent elements are the only ones associated with a unique status, then errors should be equally distributed across the remaining elements within the same utterance.

To summarize, previous studies on the distribution of speech errors have found that contextual errors are distributed unevenly within various linguistic units. Although the uneven distribution of errors is usually discussed in categorical terms – initial position is privileged or prominent elements are protected—some models that rely on activation levels might predict a more gradient distribution of errors (e.g., the PDP model reviewed above). Moreover, the data presented in Croot et al. (2010) would seem to suggest that a gradient distribution of errors may occur in supralexical prosodic units.

A gradient distribution of errors across a supralexical prosodic unit could also suggest the absence of any fixed internal structure beyond relative position. For example, a gradient decrease or increase in errors could result if activation increases or decreases as phonological encoding precedes from left-to-right (as it were) in the supralexical prosodic unit. The suggestion of increasing activation assumes an architecture analogous to the one responsible for contextual activation in the PDP model. If activation is cumulative and contextually-dependent, then this should lead to fewer and fewer errors in encoding or execution as the processes bear down on the right-edge of the unit. By contrast, the suggestion of decreasing activation assumes that the supralexical prosodic unit itself may receive activation and that this activation decays naturally and gradually over time.

A gradual decrease in activation across a planning domain could explain Croot et al.'s (2010) finding that a disproportionate number of errors occurred late in their 4-word utterances, that is, an increase in errors. Note that this contrasts with the scenario for words, which predicts a gradual increase in activation across the unit and so a decrease in errors. If we find that errors are in fact distributed differently at the lexical and supralexic levels, then this could suggest that the linguistic units involved (i.e., lexical words and prosodic feet vs. prosodic phrases or utterances) play different roles in the speech planning and production process.

2.4. Prosodic phrases as speech processing domains

Most models of speech production focus on planning and production at the word-level rather than at the sentence-level (but see Levelt, 1989 for ideas about how to model connected speech processes). Furthermore, although some models investigate the role of supralexic units in language and/or speech production (e.g., Garrett, 1975, 1980a; Levelt, 1989; Bock & Levelt, 1994), these units have typically been syntactically defined rather than prosodically defined. In spite of the focus on word-level processes in speech production, there is considerable evidence to suggest that supralexic prosodic units are involved in planning. Before reviewing this evidence, let us take a moment to define the nature and general characteristics of different supralexic prosodic units.

Whether a theory defines prosodic units in relation to syntactic structures (e.g., Selkirk, 1984, 1986; Nespor & Vogel, 1986; Hayes, 1989) or according to their suprasegmental features (e.g., Beckman & Pierrehumbert, 1986), all agree that prosodic units are hierarchically structured (see Shattuck-Hufnagel & Turk, 1996 for an overview).

Most theories postulate the Intonational Phrase (hereafter IP) as the highest units in the prosodic hierarchy, though some assume that these are embedded in an Utterance (e.g., Nespor & Vogel, 1986; Hayes, 1989; Ladd, 2008). Across theories, the IP is defined by a complete intonational contour with well-defined juncture cues such as pre-boundary lengthening and pauses (Selkirk, 1984; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992).

Below the level of the IP, theories disagree as to the number and nature of mid-level supralexical prosodic units. Some theories postulate two mid-level prosodic phrases. For example, Selkirk (1995) proposes a Major Phrase and Minor Phrase defined by syntactic phrases (e.g., Major Phrase boundaries align with the right edge of syntactic maximal projection), and Beckman and Pierrehumbert (1986) propose an Intermediate Phrase and Accentual Phrase, which are defined intonationally. Others posit just a single mid-level unit. For example, Nespor and Vogel (1986) propose the Phonological Phrase, which is defined as the domain of phonological rule application (e.g., stress shift: *thirtéen mén* becomes *thirteen mén*). Perhaps one reason for the disagreement between theorists is that mid-level units are not associated with juncture cues in the same way as the IP, at least in English. For example, according to Beckman and Pierrehumbert, the Intermediate Phrase (hereafter ip) has a nuclear pitch accent and a phrasal accent, but less obvious boundary cues. The other mid-level constituent, nested within the ip, is the Accentual Phrase (hereafter AP). These units are well described and easy to identify in French (Jun & Fougeron, 1995, 2000), Japanese (Beckman & Pierrehumbert, 1986) and Korean (Jun, 1998), but again not in English (Beckman & Pierrehumbert, 1986).

Although the exact nature and characteristics of particular units in the prosodic hierarchy are still in dispute, phonetic evidence supports the idea that supralexical prosodic units play a role in speech planning and production. One piece of evidence is that of domain-initial strengthening. Acoustic and articulatory studies have shown that initial segments are longer in duration and executed with more extreme movements of the articulators than medial or final segments, and these effects are stronger at the edges of larger prosodic units than at the edges of smaller ones (Pierrehumbert & Talkin, 1992; Jun, 1993; Fougeron & Keating, 1997; Cho & Keating, 2001; Keating, Cho, Fougeron, & Hsu, 2003; Cho, 2005, 2006). For example, Fougeron and Keating (1997) examined linguopalatal contact and duration of consonants and vowels in different positions (initial, medial, and final) within different prosodic domains (Utterances, Intonational Phrases, Phonological Phrases, Words, and Syllables). They showed that domain-initial consonants had more linguopalatal contact and longer durations, especially at the edges of utterances and IPs than at the edges of Prosodic Words. The effects were absent at syllable boundaries within a word. The differences in strengthening suggest the psychological reality of different prosodic units, and especially that all speakers differentiate IPs from words or syllables.

A second piece of evidence that supports the idea that supralexical prosodic units play a role in planning and production is found in studies that have examined their effect on pause duration (Gee & Grosjean, 1983; Ferreira, 1993; Krivokapić, 2007, 2012). For example, Krivokapić (2007, 2012) investigated the effect of prosodic complexity on pre- and post-boundary pause duration. Her results indicated that pauses were significantly shorter when the post-boundary phrase was comprised of two IPs than when it was

comprised of just a single IP. Note that since overall sentence length was controlled, the IPs were shorter in phrases with 2 IPs than in phrases with a single IP. Krivokapić suggested that the effect on pause duration indicated that speakers plan speech one IP at a time. The idea is that shorter IPs require less planning time than longer IPs, thus pause duration is shorter before a sequence of shorter IPs than before a long IP. She further suggested that a subsequent IP can be readied during execution of a current IP, meaning that a pause interval is not always required for planning.

In sum, the evidence suggests that models of speech production should incorporate supralexical prosodic units. Keating and Shattuck-Hufnagel (2002) have proposed a model that does just this (see also Shattuck-Hufnagel, 1992, for the origins of this idea). Specifically, Keating and Shattuck-Hufnagel propose a prosody-first approach to planning, arguing that prosodic structure is critical for phonological and phonetic encoding. In their view, the speech planning process begins with the construction of “a skeletal default prosody” for the whole utterance (Keating & Shattuck-Hufnagel, 2002, p.139). Here, syntactically-driven prosodic phrasing and prominence is specified. This roughly constructed frame is then restructured to accommodate influences from various types of non-syntactic information, including speech rate or phrase length. Only once all this prosodic information is available to speakers are segments retrieved and serially ordered. Thus, according to this model, speakers first chunk a sentence into prosodic phrases, and then accommodate the phonological and phonetic characteristics of phonemes to fit within the already-generated prosodic frame.

In conclusion, although very few models of speech production have investigated how a whole utterance is planned and produced, there is evidence to suggest that

information about supralexicalexical prosodic units is crucial for speech production and highly relevant to the planning process. This dissertation aims to provide additional insight into the role of supralexicalexical prosodic units during speech planning and production by investigating the distribution of errors within these units.

2.5. Current dissertation

In this dissertation, we will adopt the view that different types of contextual errors and their distributions result from the relative activation of linguistic units used in the speech planning and production processes. Based on the well-documented speech error pattern of structure preservation and the word-initial effect, we will assume that the details of phonological encoding take place at the word or prosodic foot level. We will further assume that planning also involves higher-level prosodic structures, such as the AP, ip, and IP. Such an assumption leads to the prediction that speech errors will be systematically distributed within the bounds of these units. Thus, the goal of the present dissertation is to investigate the distribution of speech errors within supralexicalexical prosodic units to understand how they are represented (i.e., their psychological reality as distinct units) and what their role is in the planning process. Note that a random distribution of errors across any of the units in question will undermine the hypothesis that they are in fact relevant to planning and production.

Four experiments were completed to test the hypothesis that supralexicalexical prosodic units are relevant to speech planning and production in adult and child language as well as in English and Korean. Experiment 1 investigated the distribution of speech errors within the IP in adult English. The findings from this experiment were that errors are

gradiently distributed across the unit, and that more errors occur towards the end of these units than at their beginning. The decrease in activation level (increase in error number) might be understood in the context of working memory processes. This idea is pursued in Experiment 2 where the effect of working memory capacity on unit size and on the distribution of errors is examined in children's speech. Experiment 2 focused on children's speech because it is less automatized than adult speech (Shallice & Butterworth, 1977; Klapp, Greim, & Marshburn, 1981; Bock, 1982), and so is likely to be more influenced by non-linguistic processes. Experiment 3 sought to replicate in another language the distributional patterns associated with the English IP. The distribution of speech errors within the AP, a mid-level supralexical prosodic unit, was also investigated. The findings were that the pattern in APs is different from that of IPs in that more errors were observed in AP-initial position than in AP-final position. The distributional patterns of errors in IPs were as in English, namely, a cumulative increase in the number of errors across the unit. Experiment 4 investigated possible explanations for the AP-initial effect. Overall, the findings were taken to suggest that mid-level prosodic units, when they are well-defined as in Korean, function more like the prosodic or lexical word in English than like the IP, which functions more as a planning domain. The theoretical implications of the results are discussed throughout and revisited in a final chapter where directions for future research are also suggested.

CHAPTER III
EXPERIMENT 1: SPEECH ERRORS
IN REAL-WORD TONGUE TWISTERS
PRODUCED BY ADULT ENGLISH SPEAKERS

The work presented in this chapter was previously reported in a co-authored article published in the journal *Laboratory Phonology*: Choe, W.K. & Redford M.A. (2012). The distribution of speech errors in multi-word prosodic units. *Laboratory Phonology*, 3, 5-26.

3.1. Introduction

Experiment 1 examined the distribution of speech errors in the supralexicall prosodic units of English in order to assess their role in the speech planning and production processes. Assuming that we find evidence for the systematic distribution of errors within supralexicall prosodic units, let us consider what these patterns might look like. Figure 3.1 illustrates four possible patterns that are predicted based on patterns previously observed for syllables, lexical words, and/or prosodic feet as well as by the patterns reported in Croot et al.'s (2010) study for phrase-level units. The four patterns also contrast predictions derived from a slot-and-filler view of serial ordering, which assumes a frame with internal structure and predicts categorical effects, and the spreading activation view, which predicts gradient effects when instead of frames with internal structures only relative position information is preserved.

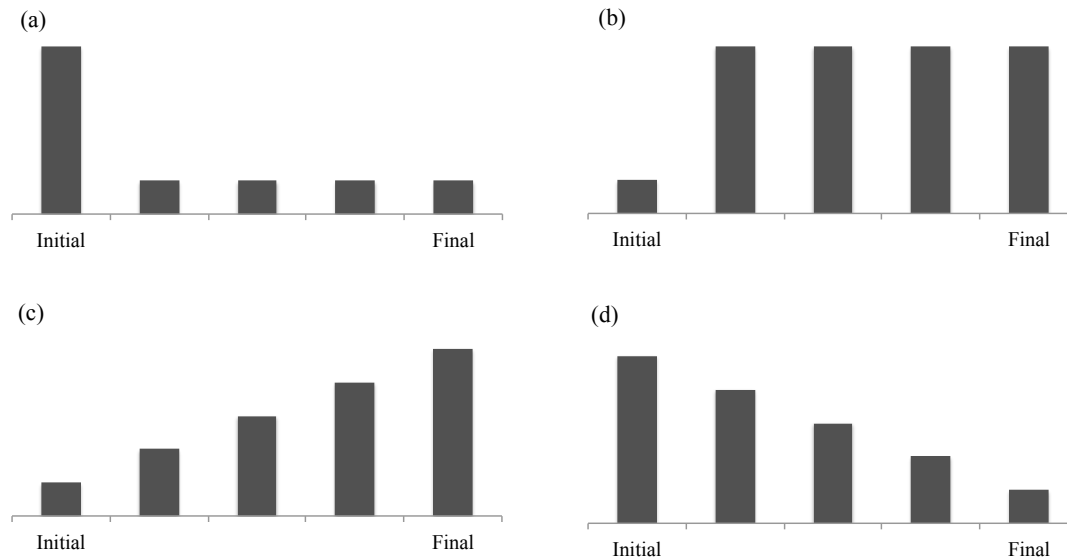


Figure 3.1. A schematic of possible patterns of error distribution within supralexic prosodic units. The horizontal axis represents the position within a supralexic prosodic unit and the vertical axis represents the number of errors. The top graphs represent categorical patterns, while the two bottom ones represent gradient patterns. See text for detailed explanation.

The first thing to notice about the patterns shown in Figure 3.1 is that they posit situations in which either the most or fewest errors occur in unit-initial position. The hypothesis that most errors will occur in unit-initial position (i.e., patterns (a) and (d)) extends patterns observed at the syllable, lexical word, or prosodic foot levels to the level of the supralexic prosodic unit (Shattuck-Hufnagel, 1987, 1992; Vousden et al., 2000; Frisch, 2000). A pattern that has a disproportionate number of errors in initial position would suggest that unit-initial elements either receives less activation than other units within the domain (e.g., Dell et al., 1993) or are treated differently from other elements within the domain (e.g., Shattuck-Hufnagel, 1987, 1992).

The alternative hypothesis shown in Figure 3.1 is that initial position is a protected position (i.e., patterns (b) and (c)). This hypothesis is based on the results from

Croot et al.'s (2010) study, where the fewest number of errors occurred on the first word in their four-word utterances. The hypothesis is also consistent with results on initial strengthening; that is, the finding that prosodic phrase-initial segments are produced with more extreme articulation and longer durations than phrase-medial and -final segments (Pierrehumbert & Talkin, 1992; Jun, 1993; Fougeron & Keating, 1997; Cho & Keating, 2001; Keating, Cho, Fougeron, & Hsu, 2003; Cho, 2005, 2006; see section 2.4 for more details). The idea is that extra articulatory effort is due to higher levels of activation, which renders phrase-initial position more resistant to error.

The next thing to notice about the patterns shown in Figure 3.1 is that the distribution of errors across other positions within the supralexical prosodic unit varies. Categorical patterns (a) and (b) are ones in which initial position differs from all other positions within the unit, and all other positions are associated either with a minimum (a) or maximum (b) number of errors. Gradient patterns (c) and (d) are ones in which the number of errors are hypothesized to vary from small to large (c) or large to small (d) in a gradient fashion across positions within the unit. The hypothesis of a categorical pattern assumes a slot-and-filler model, that errors are due to chance mis-selections, and that the initial position has a unique status in phonological encoding which is distinct from any other positions. The hypothesis of a gradient pattern assumes that the levels of activation that trigger serial ordering and the execution of elements within a unit gradually increase or decrease over the time it takes to plan and execute the specified sequence.

The current experiment tested for the possible patterns shown in Figure 3.1. Adult English speakers were asked to produce common English tongue twisters multiple times. Matching sentences were also elicited to verify that the repetition of phonological

features (a characteristic of tongue twisters) did not affect the speaker's natural prosody. Once elicited, speech errors were identified and the sentences containing them were prosodified. As noted in Chapter II (section 2.4 for details), there is disagreement as to the number and nature of mid-level prosodic units, especially in English. Beckman and Pierrehumbert (1986) suggest the Intermediate and Accentual Phrases, Nespor and Vogel (1986) the Phonological Phrases, and Selkirk (1995) the Major and Minor Phrases. In contrast, theorists generally agree on the higher-level prosodic unit, the Intonational Phrase (IP). Whereas boundary characteristics associated with IPs are highly salient, those associated with the mid-level prosodic units are much less so. It is perhaps the less salient boundary characteristics that explain the disagreement between theories as to the nature of the mid-level prosodic units in English.

Given the definitional difficulties, prosodification in the present experiment prioritized perceptual coherence. Supralexical prosodic units were defined according to that coherence and according to the juncture cues that disrupted the coherence. Whereas strong juncture cues were assumed to delimit IPs, weaker cues might delimit either an ip or an IP. In light of this ambiguity, we adopt the theory-neutral term Intonation Unit (IU) in this chapter. The core analyses investigated the distribution of speech errors within these units.

3.2. Methods

3.2.1. Participants

Forty native English-speaking University of Oregon undergraduates (19 males and 21 females) participated in this experiment. All participants received course credit for their participation.

3.2.2. Stimuli

The stimuli were 60 different sentences that varied in length such that there were 20 short sentences (4 to 7 syllables; $M = 4.50$ words), 20 medium sentences (8 to 14 syllables; $M = 6.65$ words), and 20 long sentences (15 to 25 syllables; $M = 14.10$ words). All 60 stimulus sentences are listed in Appendix A. The length of sentences was manipulated to encourage the natural prosodification of the stimuli into different sized IUs.

The sentences also varied in the extent to which phonological features were repeated: 30 sentences were tongue twisters, and 30 were not. Tongue twisters were used in order to induce sufficient errors for a meaningful distributional analysis. Non-tongue twisters were included to assess whether the error patterns elicited using tongue twisters might be similar to those which occurred in more typical speech.

Participants repeated 60 different sentences 5 times. Specifically, a total of 300 sentences were presented to participants in one of 3 pre-determined randomized orders.

3.2.3. Procedure

A single participant was seated in a sound-attenuated booth in front of a computer monitor and a standing microphone (Shure patented SM10A). The stimulus sentences

were presented orthographically on the monitor, first in red and then in green. The participant was asked to prepare the entire sentence by reading it silently when it was presented in red on the screen. Once the participant felt prepared to speak, s/he was told to press a button at which point a beep sounded and the entire sentence turned from red to green. Both auditory and visual cues signaled that the participant should start speaking. The purpose of this procedure was to maximize the likelihood that participants would construct a well-structured plan to guide speech output rather than reading the sentences word for word. Preparation times (from the time when a sentence was in red to the time when it turned to be in green) suggested that participants did in fact plan their output: a repeated measures ANOVA on the average preparation times for each speaker indicated that sentence length significantly affected planning times, $F(2,78) = 58.59, p < .01$. Post-hoc mean comparisons showed that speakers planned long sentences for longer than medium length sentences, mean difference = 603.26, $SE = 81.12, p < .01$, which they planned for longer than short ones, mean difference = 395.96, $SE = 53.22, p < .01$.

Participants were instructed to speak at a comfortable rate. They were also instructed not to stop or self-correct while they were speaking. When these instructions were not followed, participants were asked to restart the sentence. The goal of these instructions was to maximize the naturalness of speakers' prosody while also maximizing speech fluency as well as the number of errors produced. An experimenter remained with the participant throughout the session to provide feedback about speech rate and speaking level as well as to remind participants to speak as fluently as possible regardless of what errors they made.

Participants were given 10 trials to become familiar with the procedure before the test stimuli were presented. The entire session took about 45 minutes to an hour and it was digitally recorded using Marantz PMD670 for later analysis.

3.2.4. Coding

Two judges (an undergraduate research assistant and the author) listened to a total of 12,000 sentences (60 stimulus sentences \times 5 repetition \times 40 speakers) and independently identified sentences that contain speech errors⁸. Mispronunciations and disfluencies were also noted, but sentences that contained only these were not considered for further analysis. Inter-judge agreement was 93.94%. Disagreements were settled according to the opinion of the third judge, an expert in phonetics, who listened to the 1,397 sentences that were identified as containing one or more speech errors.

Next, the third judge and I independently categorized speech errors as errors of anticipation, perseveration, anticipation and perseveration, lexical substitution, insertion, or deletion. The error category depended on the presence and location of a source for the error within the stimulus sentence or preceding 2 or 3 stimulus sentences. Errors were categorized as anticipatory when the closest possible source for the error was subsequent to the error (e.g., *Greep grapes are great* instead of *Greek grapes are great*). Errors were categorized as perseveratory when the closest possible source for the error preceded the error (e.g., *Fred flew three free throws* instead of *Fred flew three free throws*). Errors were categorized as both anticipatory and perseveratory when the subsequent and preceding material could have provided the source and both were equally close to the

⁸ In the tongue twister *If Peter Piper picked a peck of pickled peppers, where's the peck of pickled peppers Peter Piper pickled*, it was not counted as an error if speaker produced the more familiar "picked" instead of "pickled" as the final word.

error (e.g., *Friendly Frank flips fline flapjacks* instead of *Friendly Frank flips fine flapjacks*). When no possible source for the error could be identified within the sentence or from preceding set of sentences, errors were categorized as errors of lexical substitution (e.g., *One smart fellow, he was smart* instead of *One smart fellow, he felt smart*), deletion (e.g., *Three grey geese in the _____ grass grazing* instead of *Three grey geese in the green grass grazing*), or insertion (e.g., *Glenwood makes fine apple flapjacks* instead of *Glenwood makes apple flapjacks*) depending on the type of error. The two coders were in agreement on 94.72% of the categorizations. Disagreements were resolved through repeated listening and discussion.

After all of the speech errors had been identified, the two coders identified prosodic boundaries. Boundary identification was originally limited to those sentences with contextual errors: namely, errors of anticipation, perseveration, and both anticipation and perseveration ($N = 966$).

The boundary identification procedure was honed on the data from 16 participants. In addition to coding the location of boundaries, two levels of prosodic boundary strength were also coded: strong and weak. Perception of boundary location and strength was repeatedly checked against spectrograms and F0 traces of the sentences. These spot checks confirmed boundary strength that strong boundaries were associated with final lengthening, a boundary tone, and a pause (longer than 200 ms). Weak boundaries were mostly identified when a pause was absent, but a boundary tone and some degree of final lengthening were present. Speech disfluencies due to self-correction or hesitation were coded separately and so are not confounded with the boundaries that are the focus of this report.

Whereas the coders jointly located strong and weak boundaries for 16 speakers, boundary location and strength for the other 24 participants' data were independently identified. Inter-coder agreement on location and strength for these data was 94.43%. Disagreements were resolved through repeated listening and by checking the visual representations of the sentences.

Once boundaries were identified in the sentences with speech errors, the agreed-upon criteria were used to identify prosodic boundaries in a matched sample of sentences without speech errors ($N = 966$). The goal of this exercise was to determine whether or not the prosodic structures of sentences with speech errors were similar to those of sentences without speech errors. Sentences without speech errors were selected speaker by speaker to match in length and phonological repetitiveness with the sentences that individual speakers produced with speech errors. These sentences were then transcribed for boundary location and strength. To ensure transcription reliability, my adviser independently coded boundary location and strength in a randomly selected subsample of the sentences without errors ($N = 250$). Inter-coder reliability was 95.20%. Disagreements were resolved in the same way as the sentences with speech errors.

3.3. Results and discussion

3.3.1. Speech errors

On average, participants produced speech errors in 8.07% of the stimulus sentences, though error rate ranged from 1% to 17.33% across participants ($SD = 4.04\%$). Altogether a total of 1,634 errors were identified in 1,397 sentences. The majority of them occurred in tongue twister sentences (77.52%). Of 1,634 speech errors, contextual

errors (i.e., errors of anticipation and/or perseveration) were the most frequent types ($N = 1,257$): 681 were identified as errors of anticipation, 507 were identified as errors of perseveration, 69 were identified as errors of both anticipation and perseveration, and 377 were identified as other types of errors (i.e., lexical substitution, deletion, or insertion). Since, by definition, only errors of anticipation and/or perseveration are contextual errors, and since these constituted the majority of all errors, the remaining analyses will focus on these types of errors.

The number of contextual errors that participants produced varied with sentence length. The fewest contextual errors were produced in short sentences ($N = 205$), more were produced in medium length sentences ($N = 463$), and the most were produced in long sentences ($N = 589$). Next, we turn to the results from the independent coding of prosodic boundaries in the 966 sentences that contained the 1,257 contextual errors and those in the matched set of 966 sentences that contained no errors.

3.3.2. Prosodic boundaries

Table 3.1 shows the number of sentences with strong and weak boundaries as a function of sentence length and the existence of speech errors in the sentences. Not surprisingly, a greater percentage of long sentences compared to medium or short sentences had at least one sentence-internal strong or weak boundary (77.11% in long sentences versus 31.52% in medium sentences and 21.16% in short sentences). The table also indicates that the distribution of boundaries in sentences with and without errors was similar across all sentence lengths and boundary types. However, there were more internal prosodic boundaries in sentences with speech errors than in those without speech

errors. This difference was significant in a chi-square test, $\chi^2(1, N = 1932) = 46.61, p < .01$.

Table 3.1. Sentences with different types of internal prosodic boundaries as a function of sentence length and whether the sentence was produced with or without speech errors.

Sentence	No internal	1 or more	1 or more
Sentence with error			
Short	115	2	46
Medium	214	7	130
Long	67	97	362
Total	396	106	538
Sentence without error			
Short	143	2	22
Medium	264	3	83
Long	139	70	295
Total	546	75	400

Note. The total number of sentences represented in the table sums to more than 1,932 (966 X 2) because a single sentence could be produced with both a strong and a weak sentence-internal prosodic boundaries.

We will refer to the units defined by prosodic boundaries as Intonational Units (IUs). The mean length of these units also varied with sentence length. Of course, short sentences with no internal boundaries were shorter than medium sentences with no internal boundaries ($M = 4.65$ words, $SD = 0.74$ vs. $M = 6.96$ words, $SD = 1.24$), which were in turn shorter than long sentences with no internal boundaries ($M = 6.96$ words, $SD = 1.24$ vs. $M = 11.18$ words, $SD = 3.11$). This pattern also held for the sentence-internal IUs defined by either strong or weak boundaries. Such units were shortest in short sentences ($M = 2.54$ words, $SD = 1.06$), longer in medium sentences ($M = 3.08$ words, SD

= 1.42), and longest in long sentences ($M = 5.15$ words, $SD = 2.37$). Mann-Whitney U tests comparing these unit lengths in sentences with and without errors indicated that these were only significantly different in medium length sentences, where unit lengths were an average of 0.52 words shorter in the sentences with errors than in those without errors, $z = -4.1$, $p < .01$.

3.3.3. Speech errors and prosodic boundaries

Scatterplots were used to evaluate the distribution of errors across positions in units of varying lengths. Position was defined as the position of the word within a unit that contained a speech error, counting from the beginning of a unit. Unit length was defined by number of words. Three types of units were considered: utterances, big IUs and small IUs. Utterances were defined by a single prosodic boundary, namely, the boundary that coincided with the sentence boundary. Big IUs were defined by a sentence-internal strong boundary and either another sentence-internal strong boundary or the beginning or ending of the sentence. A small IU was defined by a sentence-internal weak boundary on one side and either a strong or a weak boundary on the other side. Note that this definition means that big IU and small IU refer only to boundary strength, not to the size or length of an IU.

Regression lines were plotted for each scatterplot. These lines indicated a positive relationship between the position of an error within a unit and the number of words within that unit, as shown in Figure 3.2. This relationship was stronger for big and small IUs than for utterance (error-position/word-number line: utterance, slope = 0.52, $R^2 = 0.31$; big IU, slope = 0.81, $R^2 = 0.48$; small IU, slope = 0.77, $R^2 = 0.70$).

If errors had been distributed randomly across a unit, then the slope of the error-position by word-number regression line would have been 0.5. Instead, the steeper slopes indicate that the fewest errors occurred in initial position and the greatest number of errors occurred towards the end of a unit. This positional effect was much stronger in big and small IUs than in utterances, where the slope was only barely above 0.5. Then again, the shallower slope in utterances was likely due to the especially wide distribution of errors in the longest units (see Figure 3.2). When the slope was recalculated for utterances with 12 or fewer words, which was the maximum number of words in big and small IUs, it became much steeper (slope = 0.66, $R^2 = 0.36$).

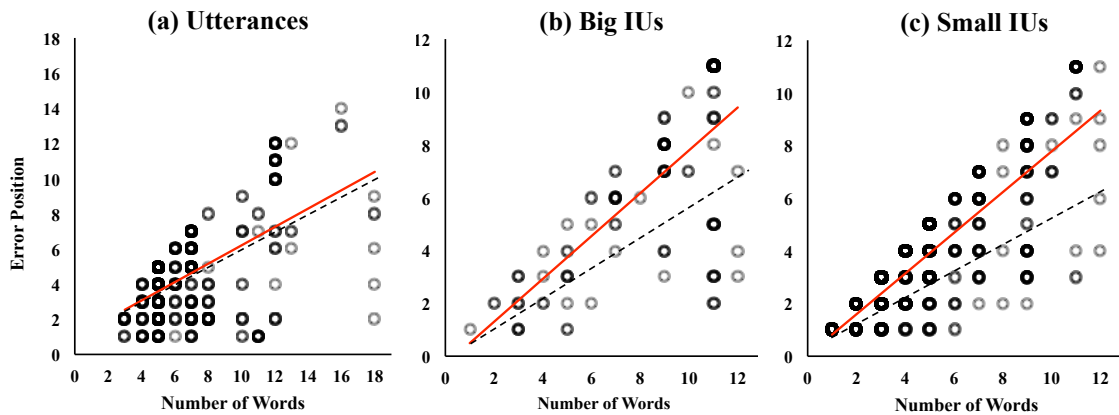


Figure 3.2. Position of words containing speech errors (error position) as a function of the number of words in utterances, big IUs, and small IUs. The circles in the figure represent errors. Darker circles correspond to greater numbers of errors for that particular position and unit length than lighter circles. The red solid lines represent actual error-position by word-number regression line and the black dashed lines represent error-position by word-number regression line assuming that errors are randomly distributed across a unit.

The especially steep slopes of the error-position by word-number regression line for big and small IUs and the high proportion of variance accounted for by these lines suggest that the positional effect was related to the IU rather than to the sentence as a

whole. To confirm that this was indeed the case, the data was divided into units that occurred early or in the middle of a sentence (sentence-nonfinal IUs) and those that occurred at the end of a sentence (sentence-final IUs). The effect of number of words within a unit on position of error within that unit was reanalyzed. The data from big and small IUs were combined for further analyses, since the previous analyses had indicated that the positional effect was similar for both types of units.

The scatterplots for sentence-nonfinal ($N = 447$) and sentence-final IUs ($N = 298$) are shown in Figure 3.3. In both cases, fewer errors occurred at the beginning of the unit than towards the end of the unit: the error-position by word-number regression line had a slope of 0.82 ($R^2 = 0.71$) for sentence-nonfinal IUs and a slope of 0.77 ($R^2 = 0.74$) for sentence-final IUs. The fact that the positional effect was equally strong for sentence-nonfinal and sentence-final IUs confirms the suggestion that the distribution of errors is related to the IU and not to sentence boundaries.

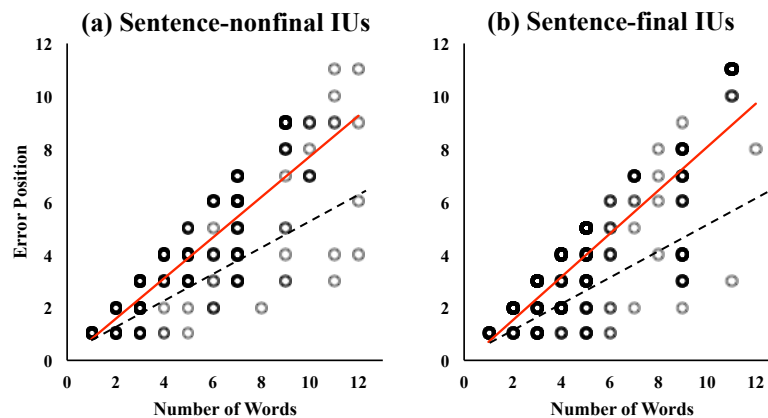


Figure 3.3. Position of words containing speech errors (error position) as a function of the number of words in sentence-nonfinal and sentence-final IUs. The circles in the figure represent errors. Darker circles correspond to greater numbers of errors for that particular position and unit length than lighter circles. The red solid lines represent actual error-position by word-number regression line and the black dashed lines represent error-position by word-number regression line assuming that errors are randomly distributed across a unit.

The next analysis examined whether the positional effect was due to a gradual increase in errors across the unit (i.e., gradient error distribution) or simply to a disproportionate number of errors in unit-final position compared to unit-initial position (i.e., categorical error distribution). Accordingly, the number of errors in 4 different position categories (initial, early, late and final position) was compared for sentence-nonfinal and sentence-final IUs of 4 or more words. Initial position was defined as the first word in the IU, early position as words in the first half of the IU, starting with the second word; late position as the second half of the IU excluding the last word; and final position as the last word in the IU. A finer grain count was also made of just IUs with odd numbers of words (5, 7, etc.). This count investigated the distribution of speech errors across 5 position categories, including a “middle” category. The patterns were found to exactly parallel those based on the previously described 4 position categories: initial, early, late, final. For this reason, only the results from the overall count are reported below.

The counts by position are shown in Figure 3.4. This distribution of errors suggests a cumulative effect. The cumulative error pattern also held for utterances: there were 38 errors in initial position, 121 in early position, 196 in late position, and 145 in final position. That is, the majority of errors (68.2%) occurred in the latter half of the utterance. To investigate whether the pattern also held for units of less than 4 words, the number of unit-initial and unit-final errors was compared for sentence-nonfinal and sentence-final IUs with 2 and 3 words. The results indicated that even in these smaller units, errors were much more likely to occur towards the end of the unit than the

beginning (sentence-nonfinal IUs, initial vs. final position, $N = 28$ vs. $N = 82$; sentence-final IUs, initial vs. final position, $N = 21$ vs. $N = 40$).

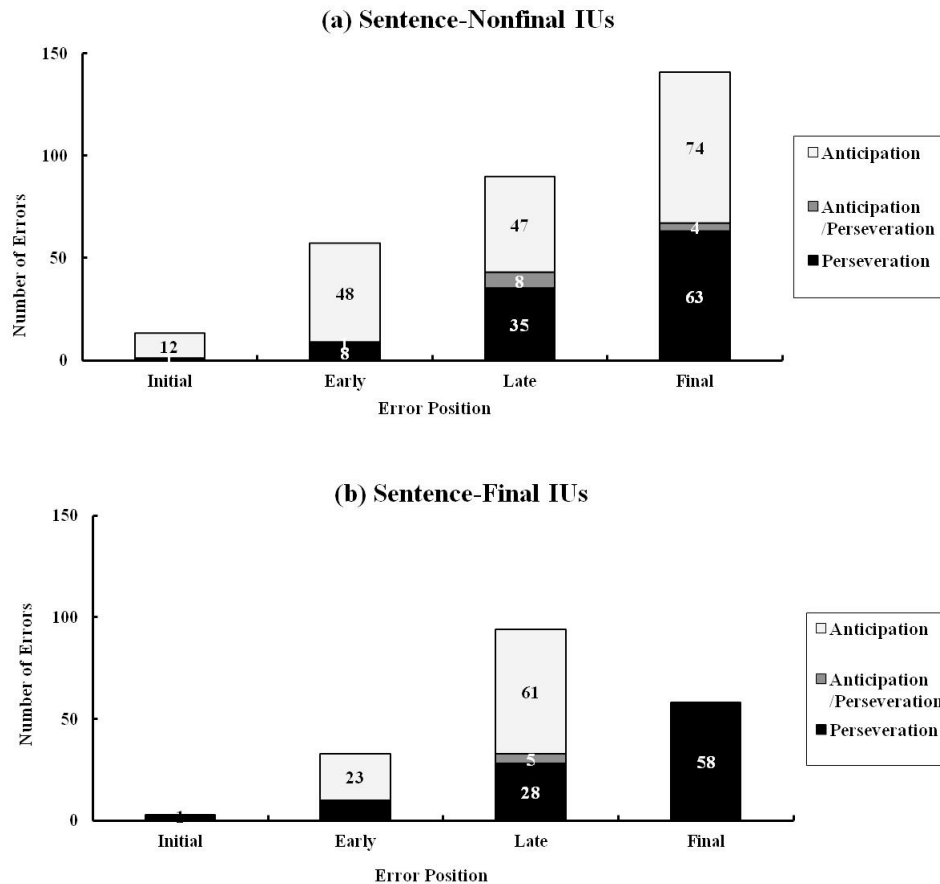


Figure 3.4. Number of errors in medium and long IUs (4+ words) as a function of the position of the words containing speech errors (error position). The proportion of errors of anticipation and/or perseveration is shown for each position as well.

Figure 3.4 also shows that the position of peak errors differed depending on whether the IU occurred within or at the end of a sentence. Specifically, most errors occurred in final position for sentence-nonfinal IUs, but in late position for sentence-final IUs and utterances. This difference suggests that the large number of unit-final errors in sentence-nonfinal IUs may be due to something other than a pattern of cumulative error. That is, additional final errors in sentence-nonfinal IUs may represent interference from

elements that occur on the other side of the sentence-internal boundary. This possibility seems especially likely when one looks at the proportion of unit-final errors of anticipation versus perseveration in sentence-final and sentence-nonfinal IUs in Figure 3.4. In particular, nearly half of the unit-final errors in sentence-nonfinal IUs were errors of anticipation ($N = 95$) and a smaller proportion were errors of anticipation and perseveration ($N = 18$), whereas all of the unit-final errors in sentence-final IUs were errors of perseveration ($N = 90$).

By definition, anticipatory and perseveratory errors take their source from a subsequent or prior element in the sentence. In the present data set, only 6% of these sequential errors took their source from within the same word. The majority took their source from either an adjacent word or a non-adjacent word within the sentence (57% and 37% respectively). This means that unit-final errors of anticipation in sentence-nonfinal IUs almost always took their source from a word on the other side of the sentence-internal boundary.

The same distribution of unit-final errors of anticipation and perseveration occurred in IUs with less than 4 words. Half of unit-final errors in smaller sentence-nonfinal IUs were due to interference from elements in the subsequent IUs (unit-final errors in sentence-nonfinal IUs, anticipatory vs. perseveratory errors $N = 33$ vs. $N = 35$; unit-final errors in sentence-final IUs, anticipatory vs. perseveratory errors $N = 0$ vs. $N = 40$).

A final set of analyses investigated whether the overall results were qualitatively similar across speakers and sentence types. Since the number of errors per speaker was very small, the number of errors that a speaker produced in the first half of an IU was

compared to the number of errors that s/he produced in the latter half of the IU. This comparison indicated that 37 out of 40 participants produced more errors in the latter half of an IU compared to the initial half, consistent with the cumulative error pattern and final position effect. Of the 3 participants who did not show this pattern, one had no contextual errors, another had only 2 errors (one in an early IU position and one in a final IU position), and the third had 8 errors (1 in initial-IU position, 4 in early position, 2 in late position, and 1 in final position).

The error distribution in non-tongue twister sentences was qualitatively similar to that of tongue twister sentences. Specifically, and in spite of the small number of total errors ($N = 56$), a cumulative error pattern emerged: there were 9 errors in initial position, 9 in early position, 23 in late position, and 15 in final position. The distributional error patterns were also qualitatively similar to those described for tongue twisters when errors in non-tongue twister sentences were investigated in sentence-nonfinal and sentence-final IUs, even though such splitting further reduced the overall number of errors to be investigated. Specifically, there was a final peak in errors for sentence-nonfinal IUs (8 errors in initial position, 4 in early, 6 in late, and 14 in final) and a late peak in errors for sentence-final IUs (1 error in initial, 5 in early, 17 in late and 1 in final position).

3.4. General discussion

The findings from this experiment indicate that the distributional pattern of contextual errors corresponds well with the boundaries of an IU, whether this unit is defined by weak or strong prosodic boundaries. In particular, the number of contextual errors was found to vary as a function of position in a gradient fashion. The obtained

pattern corresponds with pattern (c) of Figure 3.1, which we will continue to refer to as the cumulative error pattern.

The experiment also revealed an effect of IU position within the sentence on the distribution of speech errors. When IUs occurred in sentence-initial or sentence-internal position (i.e., sentence-nonfinal IUs), a disproportionate number of errors occurred in IU final position. On the other hand, when the IU occurred in sentence-final position, the number of errors associated with IU-final position was somewhat reduced relative to the number of errors in late IU position. The unique peak in IU-final position for sentence-nonfinal IUs was referred to as the final position effect. The fact that the final position effect varied with the IU position in a sentence suggests that the cumulative error pattern and the final position effect were distinct.

The reason that final position was more error prone in sentence-internal IUs than in sentence-final IUs was because a large proportion of final errors in sentence-internal IUs were anticipatory. These anticipatory errors almost always took their source from a word on the other side of the sentence-internal boundary. In this way, the final effect highlights another finding from the present experiment, namely, the absence of structure preservation in errors at the level of the IU.

3.4.1. Big and small IUs: one unit or two?

Two levels of prosodic boundary strengths (strong vs. weak) were coded in the present experiment under the assumption that the different types of boundaries might delimit different kinds of units. However, the distributional patterns of errors across the

prosodically defined units were so similar that the data from the so-called big and small IUs were combined in subsequent analyses.

The similarity between big and small IUs could mean that the distinction between different kinds of supralexical prosodic units is not relevant to speech planning or that the big and small IU represented the same unit encoded under different conditions. For example, it is possible that strong prosodic boundaries are perceived when encoding of a subsequent IP is delayed until the current IP has been fully encoded, and that weak boundaries are perceived when encoding begins in a subsequent IP before it has ended in the current one. This possibility is especially plausible if we consider the definition of big and small IUs used in this experiment. Both types of IUs were defined by juncture cues such as boundary tones and pre-boundary lengthening. Juncture in big IUs was also defined by pausing. Boundary cues such as these are a feature of IPs. In contrast, ips are defined in the literature mainly in terms of intonational cues, specifically, by a pitch accent and a phrasal accent (Beckman & Pierrehumbert, 1986). There is little to no emphasis on boundary cues. Thus, by focusing on boundary cues, it is likely that we were always coding a single type of supralexical prosodic unit, namely, the IP.

Of course, our inadvertent focus on a single unit in the present experiment should not suggest the absence of nested supralexical prosodic structures in speech planning. Instead it merely highlights the difficulty of defining these for English. In order to investigate the specific role of larger and smaller supralexical prosodic units in speech planning, it is preferable to investigate the distribution of errors in a language other than English where the different sized prosodic units can be precisely defined. This is done in

Chapters V and VI, which focus on error patterns in Korean IPs and Accentual Phrases (APs).

3.4.2. The IP as a speech planning domain

The finding that contextual errors were systematically distributed throughout IUs in the present experiment provides evidence for the psychological reality of IPs in speech planning and production. The specific distributional patterns provide us with information on serial ordering and the execution of phonemes within this domain. In particular, as an example [...*in a shoe shine sop*]_{IP} [*where she sits...*]_{IP} for *shoe shine shop* illustrates, we found that errors and their sources did not share the same IP position. This absence of structure preserving errors and the gradient error pattern suggest that the unit has little in the way of internal structure. For this reason, we conclude that the IP functions as a planning domain rather than as a frame for planning processes.

The IP delimits a chunk of the message that is generated at the morpho-syntactic level. Let us assume that this chunk, equivalent to the IP, is activated as a whole for phonological encoding. Let us further assume that IP activation interacts with the serial activation of elements (i.e., word, syllables, or phonemes) in its domain. As IP activation decays, the degree to which elements are differentially activated within the domain may also be diminished. Diminished differential activation of serially planned elements could affect sequencing in speech, resulting in an increase of contextual errors. In this way, low error rates on initial words within an IP could reflect a state of high activation, while higher error rates on subsequently occurring words would reflect the comparably lower state of overall IP activation.

Whereas the cumulative error pattern reflects decaying activation in the IP that is being encoded, the final position effect likely reflects that a subsequent chunk in the morpho-syntactic plan is being readied for phonological encoding (i.e., nearing activation threshold). That is, the final position effect can be understood to suggest that the serial production mechanism that activates the present, inhibits the past, and prepares the future (to paraphrase Dell et al., 1997), operates not only at the level of units within an IP, but also at the level of the IP itself: final words in sentence-nonfinal IPs are especially error prone because their activation is confounded by the lowered level of activation within the domain of the current IP and by the rising activation of a subsequent IP-sized sequence of words.

In summary, the cumulative error pattern and the final position effect provide evidence for the IP as a planning domain for phonological encoding in a hierarchically structured speech plan, which also clearly includes frames, albeit at the prosodic or lexical word level. The cumulative error pattern and final position effect also suggest that the serial ordering of speech depends on the gradient activation of elements within the IP domain, which preserves information about relative position.

CHAPTER IV

EXPERIMENT 2: SPEECH ERRORS IN
REAL-WORD TONGUE TWISTERS PRODUCED
BY SCHOOL-AGED ENGLISH-SPEAKING CHILDREN

4.1. Introduction

One of the reported differences between child and adult speech is that children make relatively more speech errors than adults (Wijnen, 1992; Vousden & Maylor, 2006; but see Warren, 1986). For example, Wijnen (1992) examined spontaneously produced speech errors in two young boys, aged 2 and 3 years old, and found that these children made 7 times more errors than adults in the London-Lund corpus. The errors were otherwise similar to those found in adult speech: the errors preserved syllable structure and there was a word-initial effect (i.e., speech errors are most likely to occur in word-initial position than in any other position within the word). Vousden and Maylor (2006) also report high error rates in older children's speech. These researchers used 4-word tongue twisters (e.g., *fine fresh free fish*) to experimentally elicit speech errors from 8 and 11 year old children as well as from young adults. They found that the younger children made 2.4 times more errors than the older children, who in turn made twice as many errors as young adults. The fact that error rates vary with age suggests that the ability to plan and produce speech change over developmental time.

In addition to higher error rates, children make relatively fewer anticipatory errors and more perseveratory errors than adults (Stemberger, 1989; Vousden & Maylor, 2006;

but see Wijnen, 1992 and Jaeger, 1992 for contradictory findings). This finding is usually expressed in terms of an anticipatory proportion (AProp; Schwartz, Saffran, Bloch, & Dell, 1994; Dell et al., 1997; Vousden & Maylor, 2006), which is calculated by dividing the number of anticipatory errors by the total number of contextual errors (anticipatory + perseveratory). Accordingly, an AProp value higher than 0.5 indicates that contextual errors are predominantly anticipatory rather than perseveratory, and vice versa for an AProp value lower than 0.5. Vousden and Maylor reported AProp values of 0.42 and 0.45 for 8-year-olds and 11-year-olds, respectively, and 0.61 for young adults.

Vousden and Maylor (2006) offer two explanations for the perseveratory-dominant error pattern in children's speech. The first references Stemberger's (1989) argument that school-aged children are less able than adults to inhibit the past, so active elements persist for longer during planning and production in children's speech. The second explanation references Dell et al.'s (1997) argument that the activation of future elements in a planned sequence increases with practice, and children have less overall speech practice than adults. The developmentally-related differences in practice means that children's plans offer less in the way of look ahead than adult plans. Although Vousden and Maylor's data are insufficient to distinguish between the two explanations, both explanations clearly link AProp values to planning and suggest, once again, that speech planning processes develop over an extended period of time.

What might explain why children as old as 11 years have immature speech planning and production abilities? One explanation might be that children have yet to acquire the high-level prosodic units that define the extended speech planning domain. This explanation seems highly unlikely, though, since IPs are readily identified in the

speech of much younger children (Snow, 1994; Behrens & Gut, 2005; Chen & Fikkert, 2007; Prieto & Vanrell, 2007). For that matter, the notion of holophrastic speech—one-word utterances produced under a complex, and sometimes meaningful, intonation contour (see Tomasello, 2003)—suggests that IP-like units are present from the outset of language production (Redford, to appear). Thus, a more likely explanation for the differences in child and adult speech errors may be the differences in the cognitive abilities underlying speech planning. As originally noted in Chapter I, one critical non-language ability necessary for speech planning is working memory.

The relationship between working memory and language is attested in child language⁹. For example, Adams and Gathercole (1995, 2000) found that children with larger working memory capacities produced longer IPs¹⁰, had more diverse vocabularies and more complex syntactic structures than children with smaller working memory capacities. Although Adams and Gathercole interpreted the relationship between working memory and language abilities to mean that children with larger working memories are better language learners than children with smaller working memories, the findings might also suggest something about planning and the changes in planning that result from developmental increases in working memory capacity. One specific idea that follows from our interpretation of the role of the IP in planning (Chapter III) is that speakers with larger working memory capacities may not only hold more material in working memory

⁹ The relationship between working memory capacity and length of utterances was not found in adults' speech production (Shallice & Butterworth, 1977; Klapp, Greim, & Marshburn, 1981). One explanation for the absence of a correlation is that phonological encoding in adults is highly automatized so the effect of working memory capacity is minimal under conditions that do not provide additional tax on working memory (Bock, 1982).

¹⁰ Adams and Gathercole use the term “utterance” rather than IP, but since their definition of an utterance was a unit with “apparent terminal intonation contour” (Adams & Gathercole, 2000, p. 104), I have equated their “utterance” to the IP.

for planning, but may also be able to sustain activation of this material for longer than those with smaller working memory capacities. In other words, the temporal decay in activation of the IP could be faster in children than in adults because children have smaller working memory capacities than adults. Note that the hypothesis of more rapidly decaying activation in children than in adults is also consistent with lower AProp values in children if these are indeed due to the lesser activation of “future” elements in a plan.

The goal of the current experiment was to investigate the relationship between working memory and the distribution of errors within various linguistic units (i.e., syllables, lexical words, and supralexical prosodic units) in children’s speech. Based on previous literature, we expected to find higher error rates and lower AProp values than in adult speech. But we also expected to find structure preservation, a word-initial effect, and a cumulative error pattern within the IP, just as in adult speech. Following the hypothesized relationship between working memory and planning within the domain of an IP, we also predicted that children with larger working memory capacities would produce longer IPs and more anticipatory errors than children with smaller working memory capacities. Either way, the results were expected to further our understanding of the nature and role of the IP in speech planning and production.

4.2. Methods

4.2.1. Participants

Thirty-four children participated in the current experiment. At the time of the experiment, children ranged in age from 6;3 to 9;10, with an average age of 8;0 years (*SD* = 9.5 months). All participants were native speakers of English, and 23 of them were

female and the rest were male. All children were developing typically according to parental report, and had normal hearing as determined by a pure-tone hearing screen. The Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007) was administered to provide a rough assessment of language development. The 34 children's raw scores ranged from 125 to 182 ($M = 152.85$; $SD = 16.75$) with standardized scores that ranged from 82 to 150 ($M = 120.03$; $SD = 14.59$), which represent average to well above average receptive vocabulary scores for a given age.

4.2.2. Standardized assessment of working memory

Verbal working memory capacity was assessed using Nonword Repetition (hereafter, NR), a subtest of the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgeson & Rashotte, 1999). In this subtest, a participant listened to 18 pre-recorded nonsense words, whose lengths varied from 1 to 6 syllables, and repeated these as clearly as possible. The 34 children's raw scores ranged from 8 to 18 ($M = 13.28$; $SD = 2.72$). Standardized scores, which are scores that control for a child's age, ranged from 7 to 16 ($M = 10.58$; $SD = 2.17$), which is equivalent to a range that spans from the 16th to 98th percentile.

A Pearson product-moment correlation coefficient was computed to assess the relationship between working memory capacity and language development. The analysis showed significant correlation between the standardized NR scores and the standardized PPVT scores, $r(32) = .52$, $p = .002$. That is, children with larger working memory capacities for their age had larger vocabularies for their age than children with smaller

working memory capacities. This result is in line with Adams and Gathercole's (1995, 2000) findings.

4.2.3. Elicitation: the tongue-twister task

The stimuli were 6 tongue twisters selected from the set of tongue twisters used in Experiment 1. Two of these were medium length tongue twisters and the rest were long. The six are marked with asterisks in Appendix A¹¹. These specific tongue twisters were selected because they had generated a substantial number of errors in adults (i.e., error rates greater than 10%), and were usually prosodified by adults into at least two IPs. The selected tongue twisters consisted of familiar words, thereby minimizing the chance that children would produce “reading errors” instead of contextual errors. One additional short tongue twister was selected for practice (*Frank flips fine flapjacks*).

The tongue-twister task was divided into three parts: reading practice, repetition practice, and repetition of target tongue twisters for error elicitation. During reading practice, the experimenter presented the child participant with a large flash card that had the target tongue twisters written in large (30 point), easy to read, font. The experimenter then asked each participant to read each tongue twister, starting with the practice tongue twister. Any difficulties in reading particular words were remedied at this stage. In particular, when a child could not read a tongue twister fluently, the experimenter either read the specific mispronounced/unfamiliar word for the child and then elicited a repetition or read the entire tongue twister in word-by-word fashion, with children

¹¹ In Experiment 2, the last word of the tongue twister, *If Peter Piper picked a peck of pickled peppers, where's the peck of pickled peppers Peter Piper pickled*, was changed into “picked”, since this is more familiar version of the tongue twister than the one used in Experiment 1.

following along. Word-by word reading was used to minimize effects of the experimenter's own intonation and phrasing patterns on the children's acquisition of the tongue twister. The goal was to make children familiar with the words and syntactic structure of the stimuli so that elicited speech errors would not be due to reading processes but to phonological planning and encoding process.

When participants were able to read the tongue twisters fluently and without aid, they moved on to practice repeating the tongue twisters. Each of the tongue twisters was repeated three times. Several specific instructions were given and practiced in order to elicit the best experimental data possible. First, participants were asked to take a big breath between each repetition in order to prevent self-interruptions due to poor breath control. Second, participants were asked to produce tongue twisters at a normal speaking rate—children naturally speed up the rate of repetition in a task like this. Lastly, participants were asked to continue speaking even if they made an error during a repetition. Participants followed these instructions with the experimenter providing continuous coaching and feedback.

Once the experimenter verified that participants were able to produce the target tongue twisters and that they understood the instructions regarding repetition, s/he shuffled the 6 flash cards (without the practice tongue twister) and asked each participant to randomly pick one for repetition. Participants then repeated the selected tongue twister 6 times with the experimenter counting the number of repetitions on his/her fingers in full view of the child participant. If a participant introduced a disfluency (stopping due to self-correction or breathing), s/he was asked to redo the repetition task for that tongue

twister. After 6 fluent repetitions, the child was asked to select another flash card and the process repeated until all tongue twisters had been repeated.

In the end, if a participant had significant difficulty in fluently repeating a particular tongue twister, errors elicited from that tongue twister were excluded from analyses. In total, 5 participants failed to render one tongue twister fluently, and 2 participants failed to render 2 tongue twisters fluently, and 1 participant failed to render 3 out of the 6 tongue twisters fluently. Data from this last child (aged 6;5) were excluded altogether from the analyses. Analyses were therefore based on a total of 1,134 sentences (6 sentences \times 6 repetitions \times 26 participants; 5 sentences \times 6 repetitions \times 5 participants; 4 sentences \times 6 repetitions \times 2 participants).

The whole procedure took about 15 to 20 minutes and participants performed the task in a quiet laboratory room. Speech was digitally recorded onto a Marantz PMD660 for later analysis using a Shure ULXS4 standard wireless receiver and lavalier microphone.

4.2.4. Coding

Two judges (an undergraduate research assistant and the author) listened to all 1,134 sentences and independently identified speech errors. Inter-judge agreement was 95.50%, and disagreements were resolved through repeated listening by the author.

The author then categorized speech errors according to the criteria used in Experiment 1. Errors of anticipation, perseveration, anticipation and perseveration, deletion, insertion, or lexical substitution were coded. In addition, each contextual error (anticipation, perseveration, anticipation and perseveration) was tagged with information

regarding syllable structure (an onset, a nucleus, and a coda) and whether this was shared with the source. Errors were also coded for syllable position within a word. This additional coding allowed for the investigation of structure preservation and word-initial effects in children's elicited speech errors.

The author prosodically transcribed all of the 1,134 sentences regardless of error occurrence. The location and the strength (strong and weak) of prosodic boundaries were coded based on the prosodic boundary identification criteria that were established in Experiment 1. Once again, 3 types of supralexical prosodic units were defined: utterances by a single prosodic boundary, that is, a boundary that coincided with the sentence boundary; big IUs by either two sentence-internal strong IP boundaries or a single sentence-internal strong IP boundary, and the beginning or end of the sentence; and small IUs, defined by a weak IP boundary on one side and a strong or a weak IP boundary on the other.

4.3. Results and discussion

4.3.1. General characteristics of children's speech errors

Out of the 1,134 sentences analyzed, 649 contained one or more speech errors (57.23% of the sentences). In individual children, errors occurred in between 30.55% to 94.44% of the sentences produced ($M = 54.08$; $SD = 13.95$), indicating that school-aged children typically made speech errors in at least half of the tongue twisters uttered. This average error rate was extremely high in comparison with that of adults' speech: the average error rate in Experiment 1 was only 8.07%. The higher average error rate in the current experiment is likely at least partially due to differences in the elicitation task (e.g.,

only tongue twisters as stimuli and repeating each tongue twister in a row). That said, the high error rate is also consistent with previously published reports of error rates in children's speech.

A total of 1,085 errors were noted and categorized in the 649 sentences with errors. Of these, 67% were contextual errors: 298 were categorized as anticipatory errors, 340 as perseveratory errors, and 89 as anticipatory and perseveratory errors. Thus, an overall perseveratory dominance was observed ($AProp = 0.47$). When $AProp$ was calculated for each child, the results indicated that the mean $AProp$ value was 0.48 ($SD = 0.13$), consistent with a perseveratory-dominant error pattern. This pattern can be compared to the anticipatory dominant pattern found in adult speech: the mean $AProp$ value for adults in Experiment 1 was 0.58 ($SD = 0.13$). Though the $AProp$ values are not strictly comparable due to differences in the elicitation tasks, they are nonetheless in-line with what is reported in previous literature (Stemberger, 1989; Vousden & Maylor, 2006) and with the idea that the phonological planning and encoding processes are not yet fully developed in school-aged children.

We turn now to the distribution of children's errors in linguistic units, beginning with syllables and lexical words. These analyses focused on the structure-preserving nature of speech errors and the word-initial effect. With respect to syllable structure, the analysis revealed that 93.12% of contextual errors shared the same syllable position as their source. With respect to the word-initial effect, the analysis showed that 60.12% of the 504 errors in monosyllabic words occurred in onset position, 27.38% occurred in the nucleus, and 12.50% in coda position. The results were similar for the 223 errors that occurred in multisyllabic words: 47.01% of the errors occurred in word onset position,

22.65% in the remainder of the first syllable, and 30.34% in subsequent syllables. Taken together these findings suggest that the planning and encoding processes in school-aged children’s speech reference syllable structure and lexical words just as in adults.

Before considering whether children’s errors are also distributed within supralexicalexical prosodic units in the same way as in adult speech, we first describe the supralexicalexical prosodic units in question. Table 4.1 shows the average number and length of big and small IUs in a single repetition of a tongue twister as a function of whether the tongue twister was produced with at least one contextual error or not. Mann-Whitney *U* tests were performed to assess whether the number and the length of big and small IUs varied systematically as a function of error occurrence. The results indicated that they did not.

Table 4.1. Mean number and length in words of big and small IUs in a single repetition of a tongue twister as a function of contextual-error occurrence.

Error occurrence	IUs	Number (SD)	Length (SD)
With error (<i>N</i> = 308)	Big IUs	0.76 (1.18)	10.29 (3.91)
	Small IUs	2.29 (1.17)	5.05 (1.39)
Without error (<i>N</i> = 321)	Big IUs	0.68 (1.07)	10.47 (3.75)
	Small IUs	2.31 (1.17)	5.09 (1.26)

Note. This table excludes tongue twisters that were produced without any internal boundaries (*N* = 202): 124 of the excluded sentences did not contain contextual errors and the rest contained contextual errors.

The distribution of errors within IUs was once again assessed by plotting the position of a word with an error (error position), counting from the beginning of a unit,

against the number of words in the IU. An error-position by word-number regression line was then inserted to ascertain the extent to which the slope deviated from 0.5, which would indicate a random distribution of errors across the unit. The plot with associated regression line is shown in Figure 4.1.

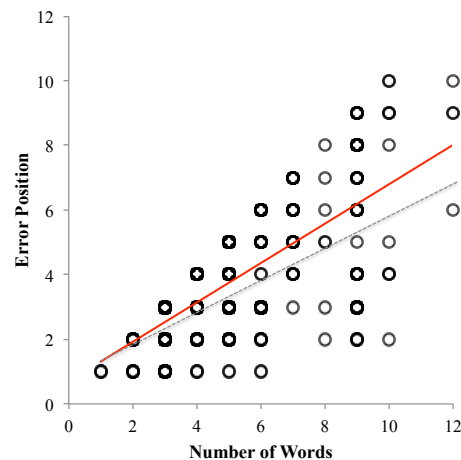


Figure 4.1. Position of words containing speech errors (error position) as a function of the number of words in the IUs. Darker circles correspond to greater numbers of errors than lighter circles. The red solid line represents the actual error-position by word-number regression line and the black dashed line represents error-position by word-number regression line assuming a random distribution of errors across the unit.

Note that the plot in Figure 4.1 combines the data from big and small IUs. This is because the analyses indicated that the slopes for big and small IUs in children’s error data were similar to each other (big IU slope = 0.63, $R^2 = 0.46$; small IU slope = 0.60, $R^2 = 0.50$), just as in adult speech. Also like adult speech, the overall slope of the error-position by word-number regression line for IUs indicated that more errors occurred towards the end of the IU than the beginning (slope = 0.61, $R^2 = 0.48$). However, in the adult data, the average slope was substantially steeper (slope = 0.79), indicating that the cumulative error pattern was weaker in the child data than in the adult data. The detailed

distributional pattern of errors within sentence-final and sentence-nonfinal IUs is shown in Figure 4.2.

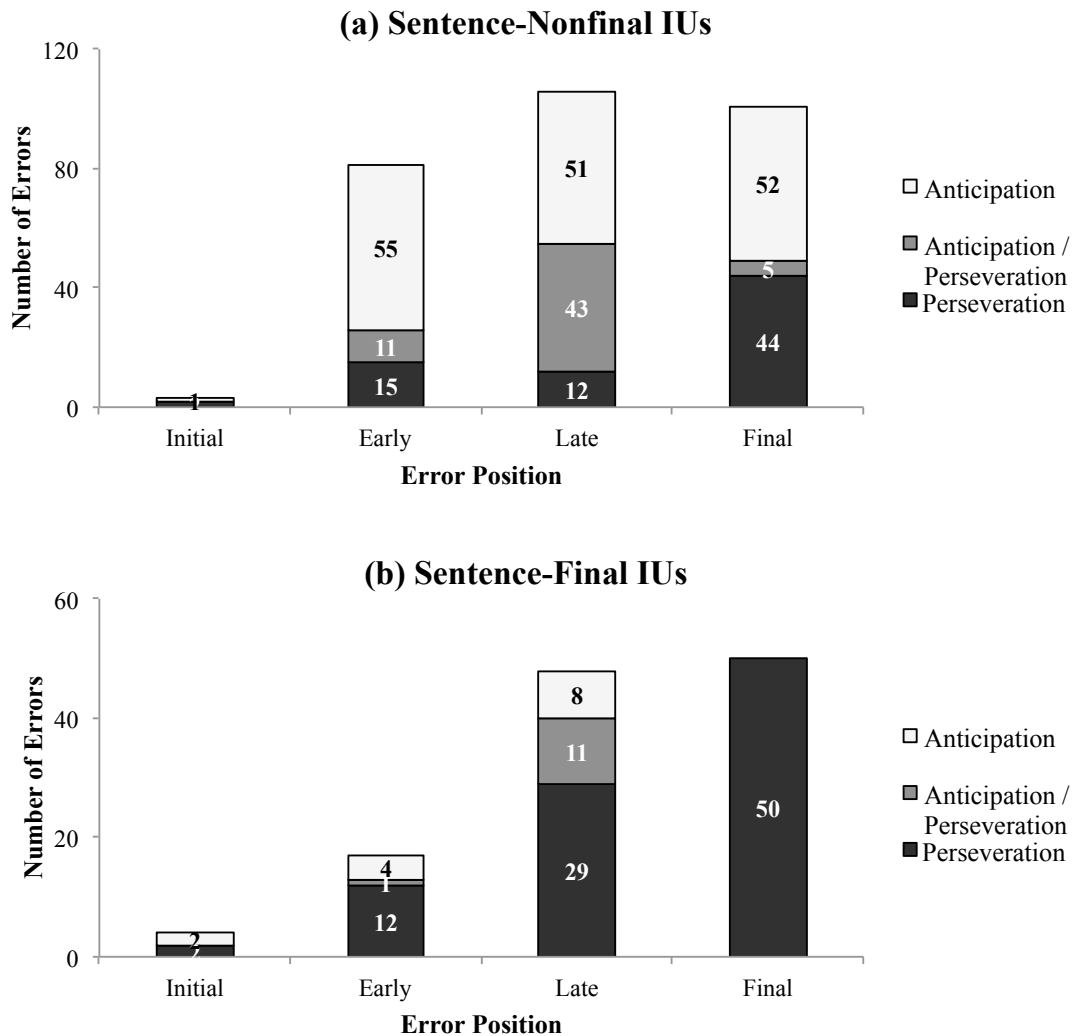


Figure 4.2. Number of errors in medium and long IUs (4+ words) as a function of the position of the words containing speech errors (error position). The proportion of errors of anticipation and/or perseveration is shown for each position as well.

Figure 4.2 provides additional information regarding the number and types of contextual errors that occurred across IU positions in children’s speech. Again, the cumulative error pattern is clear as the fewest errors occurred in initial position, more in

early-mid position, even more in late-mid and final positions¹². However, unlike the adults' error data, a final position effect (i.e., the unique peak in IU-final position for sentence-nonfinal IUs) was not observed in children's error data. Figure 4.2 also shows that the number of perseveratory relative to anticipatory errors also increases with IU position, which is similar to the adult pattern (Figure 3.4). The systematic distribution of errors within the IU and the steady decrease in the relative proportion of anticipatory error strongly suggest that the IP serves as a planning domain in children's speech just as in adults' speech.

4.3.2. The effect of working memory on speech planning

The final analyses tested the central hypothesis of a relationship between working memory and planning in children's speech. Standardized NR scores were used to predict IU length and AProp values using linear regression. The average length of error-containing IUs and the average length of all IUs was calculated across sentences within each participant. The analyses showed that standard NR scores significantly predicted the length of error-containing IUs, $\beta = -.35$, $t(31) = -2.07$, $p = .047$, but not of all IUs, $\beta = -.22$, $t(31) = -1.25$, $p = \text{NS}$. Further investigation, however, revealed a significant effect of standard NR scores on the average length of big IUs, $\beta = -.38$, $t(31) = -2.32$, $p = .027$. In either case, the relationship between working memory and IU length was opposite of the one predicted: higher working memory scores were associated with shorter, not longer, IUs (see Figure 4.3). I entertain a possible explanation for the unexpected result in the general discussion below.

¹² The same cumulative error pattern was seen in IUs with less than 4 words. In these short IUs, 23 of the 84 errors occurred in initial position and 61 in final position.

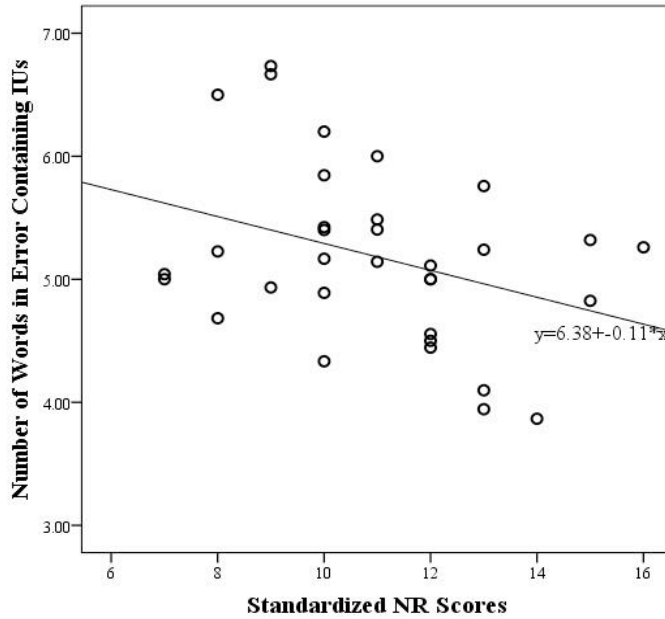


Figure 4.3. Average length in words of error-containing IUs as a function of the standardized NR scores.

The analysis on AProp values was in line with the prediction. Standard NR scores significantly predicted AProp values, $\beta = .36$, $t(31) = 2.14$, $p = .040$, such that higher working memory scores were associated with higher AProp values. The result, shown in Figure 4.4, indicates that children with larger working memory capacities produced more anticipatory contextual errors than children with smaller working memories. We discuss this expected result in the context of the unexpected result on IU length below.

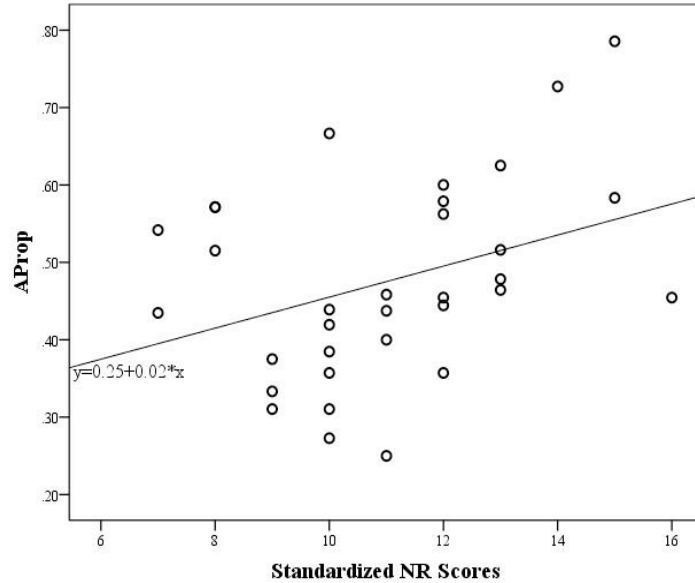


Figure 4.4. AProp values as a function of the standardized NR scores.

4.4. General discussion

The present experiment examined the characteristics of errors in school-aged children’s speech. The observed error patterns were consistent with those reported in the literature. In particular, children had higher error rates than adults and their errors were perseveratory-dominant rather than anticipatory-dominant. The distribution of errors across several linguistic units was also examined. The patterns were similar to adult patterns: the errors preserved syllable structure, there was a word-initial effect, and there was a cumulative error pattern in the supralexical prosodic units. However, the distributional pattern of errors within sentence-final IUs was similar to that of errors within sentence-nonfinal IUs. The absence of a final position effect suggests that the execution of the current IU in child speech is less affected by the plan of a subsequent IU than in adult speech. As in the adult data, there were no differences in the distribution of errors within the so-called small and big IUs. Accordingly, we will assume as before that

in spite of the differences in boundary strength, both units were in fact IPs. Altogether, these findings provide evidence that school-aged children have similar speech planning strategies and structures as adult speakers. The differences observed (higher error rates, lower AProp values, weaker cumulative error patterns, and no final position effect) suggest, however, that school-aged children are still developing their planning abilities.

The idea that planning abilities continue to develop well into middle childhood led us to predict that planning is more directly influenced by verbal working memory capacity in children than in adults, where there are no reported correlations between working memory and utterance length (Shallice & Butterworth, 1977; Klapp, et al., 1981). The specific predictions were that working memory would be positively correlated with the size of the planning domain (i.e., the IP) and with the length of time that a child could hold this plan in memory for encoding and fluent execution (i.e., AProp values). Instead the age standardized working memory scores were negatively correlated with IP length, even though they were positively correlated with AProp values. The unexpected results regarding the relationship between memory and IP length leads us to revise somewhat our understanding of the role of the IP in speech planning.

The finding that working memory and IP length are negatively correlated in children's speech is very surprising given Adams and Gathercole's (1995, 2000) finding that children with higher nonword repetition scores produce longer utterances than children with smaller working memory. Of course, their definition of an utterance was pause-delimited speech. This is not necessarily the same as an IP, which requires the presence of a prosodic boundary, but not of a pause per se (e.g., small IUs). It could be that the utterances that Adams and Gathercole examined were more equivalent to

sentences than to IPs, and so indicated something about a child's developing syntactic abilities rather than their planning abilities. Assuming that this is the case, then it would explain why we also find (like Adams and Gathercole) that verbal working memory predicts receptive vocabulary size. But, of course, we are still left with the surprising result that working memory and IP length are inversely related.

Although the result that negative correlation between working memory and IP length is surprising, the positive correlation between working memory and AProp values suggests that the results may still be relevant to planning. More specifically, the results may be specific to planning tongue twisters, which are, after all, designed to be difficult to produce.

Let us assume as before that children with better working memories are more capable of speech planning than children with poorer working memories. In the present context, better planning could amount to chunking the intended utterances into smaller domains. This kind of chunking strategy may be a way to minimize serial ordering errors, assuming that these emerge due to similarity based interference from elements that have been activated as part of the encoding process. By the same token, if the children with better working memories are also able to hold the planned domains in memory for longer than children with poorer working memories, then this would explain why children with higher NR scores also produced a higher proportion of anticipatory errors than children with lower NR scores.

In sum, the results from the present experiment suggest that children who have better working memories, defined here as larger capacities, may chunk their speech more thoroughly for planning, and may be able to hold these plans in working memory for

longer than children who have relatively poorer working memories. Thus, the results provide further evidence for our contention that the IP represents a planning domain in speech, and that the encoding processes that take place within this domain are subject to the constraints of working memory.

CHAPTER V

EXPERIMENT 3: SPEECH ERRORS

IN REAL-WORD TONGUE TWISTERS

PRODUCED BY KOREAN SPEAKERS

Part of the work presented in this chapter was previously reported in the peer-reviewed proceedings from the 17th International Congress of Phonetic Science: Choe, W.K. (2011). The distribution of speech errors in prosodic phrases in Korean. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS-11 Hong Kong)* (pp. 480-83).

5.1. Introduction

Results from the experiments presented in Chapters III and IV indicated that contextual errors are distributed in a gradient fashion within a supralexical prosodic unit with errors least likely to occur in initial position and most likely to occur towards the end of the unit. The findings were also that the cumulative error pattern was the same regardless of differences in boundary strength. Although this could mean that the different units referred to as big and small IUs are treated similarly in the speech plan, we argued that it is more likely that these were actually the same unit, namely, the Intonational Phrase (IP). The problem lies with defining intermediate and smaller supralexical prosodic units for English. English is therefore a poor language for testing the possibility that larger and smaller supralexical prosodic units represent distinct

domains in speech planning and production. Accordingly we turn now to an investigation of the distribution of speech errors in Korean, a language in which the different supralexical prosodic units are better defined and more distinct than in English.

An additional benefit of studying speech errors in Korean is that most extant studies on the distribution of speech errors have focused on Germanic languages, specifically English, Dutch, and German (Hokkanen, 2001). What little cross-linguistic data is available for non-Germanic languages suggest that certain patterns we take for granted, such as structure preservation and the word-initial effect, may be less salient in other languages (see Hokkanen, 2001 and references therein for Finnish; Chen, 1999, 2000; Chen, Chen, & Dell, 2002; O'Seaghdha & Chen, 2009; O'Seaghdha, Chen, & Chen, 2010 for Mandarin Chinese; García-Albea, del Viso, & Igoa, 1989; Berg, 1991; Pérez, Santiago, Palma, & O'Seaghdha, 2007 for Spanish). The cross-linguistic differences suggest that certain aspects of speech planning are language specific. To take a specific example, Chen (2000) reported that, in contrast to English speakers, speakers of Mandarin Chinese produce more whole syllable errors than segmental errors. Based on these data, Chen argued that the syllable is more fundamental to phonological encoding in Mandarin Chinese than the phoneme. We expected that the pattern of error distribution in Korean would provide similar language-specific insights into the speech planning and production process.

The various dialects of Korean have different prosodies. Our focus will be on the prosodic structure of the standard Korean dialect, which has been thoroughly described by Jun (1998, 2000, 2005a, 2005b, 2005c). We follow her prosodic analysis here. The standard dialect is the Seoul dialect of Korean, also known as Kyoung-gi Korean.

Standard Korean (hereafter simply Korean) is syllable-timed; that is, in contrast to English, the rhythm structure of the language is not organized around lexical stress. As for intonation, Korean has three hierarchically organized prosodic units: the Accentual Phrase (AP), the Intermediate Phrase (ip), and the Intonational Phrase (IP). Each of these units is described below.

The AP is the smallest supralexicical prosodic unit in Korean. It is tonally demarcated, defined by one of two tonal patterns: Low-High-Low-High (LHLH) or High-High-Low-High (HHLH). The exact sequence depends on the AP-initial segment: when the AP-initial segment is aspirated, tensed, /s/, or /h/, the pattern is HHLH, otherwise it is LHLH pattern. The AP is three to five syllables long on average. When the AP is shorter than four syllables, the middle falling tone (HL) can be undershot. The AP is supralexicical in that it can contain one or more lexical words, but in many cases it contains just one (Kim, 2004; Jun, 2005b). The AP is defined completely by the tone sequence. There is no obligatory final lengthening or other temporal boundary marking cues.

The ip is the next larger prosodic unit in Korean. It is the domain of focus. In her revised intonational phonology of Seoul Korean, Jun (2005a) argued that ip boundaries are marked by a pitch reset on the ip-initial syllable or by a higher than expected AP-final tone. Like the AP, the final boundary of an ip is also not obligatory marked with final lengthening. Note that the AP is nested within the IP, but not necessarily within an ip. This is because the ip emerges only under unusual circumstances (focus marking and complex syntax such as relative clauses).

As in English, the largest supralexicical prosodic unit in Korean is the IP. This unit is marked by one of nine boundary tones, which are not necessarily contrastive (e.g., Lee,

1990; Park, 2013). Since the prosodic phrases in Korean are hierarchically organized, the AP-final tone is preempted by the IP-final boundary tone. In addition to the boundary tone, obligatory final lengthening marks the right-edge of an IP. An IP is usually but not obligatorily followed by a pause.

Experiment 3 examined the distribution of speech errors within Korean prosodic phrases, but also at the lexical and syllabic levels. The obtained distributional patterns then were compared to those of English (in Experiment 1). The goals were to test the hypothesis that intermediate supralexical prosodic units are relevant to the planning and production process, and to examine the extent to which the distributional patterns associated with English IPs, words, and syllables are language-specific or also a feature of Korean.

5.2. Methods

5.2.1. Participants

Twenty native speakers of Korean participated in the current experiment (6 males and 14 females). All participants spoke either the Seoul or the Kyoung-gi dialects of Korean. All were undergraduate or graduate students at the University of Oregon, who had lived in the United States for no more than 5 years at the time of the experiment.

5.2.2. Stimuli

The stimuli were 26 different Korean tongue twisters that ranged in length from 1 to 3 sentences (1 tongue twister had 2 sentences; 1 tongue twister had 3 sentences¹³) and

¹³ Here, a sentence refers to a unit that was presented to a participant with a period or a question mark. As noted here, some of the tongue twisters contained more than one sentence.

from 7 to 150 syllables. Again, variation in length was used to ensure that participants would prosodify the sentences in a natural manner. A complete list of the tongue twisters used in the experiment is provided in Appendix B.

5.2.3. Procedure

The procedure in Experiment 3 was identical to that in Experiment 1, except that no filler sentences were included. Also each participant repeated each tongue twister 6 times so that a total of 156 tongue twister productions were elicited.

The whole procedure took between 45 minutes to an hour to complete. The experiment was performed in a quiet laboratory room. Speech was digitally recorded to a Marantz PMD660 using Shure ULXS4 standard wireless receiver and lavalier microphone.

5.2.4. Coding

Two coders (a female native speaker of Korean and the author, also a native speaker of Korean) listened to the 3,120 tongue twisters that were elicited from the 20 speakers (26 tongue twisters \times 6 repetitions for 156 tongue twisters per speaker), and independently coded these for occurrences of speech errors. Inter-coder reliability was 93.32% and disagreements were resolved through repeated listening and discussion between the coders.

The author then categorized speech errors as errors of anticipation, perseveration, anticipation and perseveration, deletion, insertion, or lexical substitution following the criteria established in Chapter III. Hesitation and disfluencies were noted but not counted

as speech errors. In addition, contextual errors and their sources were marked for segmental information (consonant vs. vowel errors) and syllable structures (onset, nucleus, and coda errors) so that the error characteristics could be examined and the distributional patterns fully evaluated.

All repetitions of tongue twisters that contained contextual errors were prosodically transcribed by the author. The boundaries of the three supralexical prosodic units were noted following the conventions laid out in the Korean Tones and Break Indices (K-ToBI; Jun, 2000) except for ip boundaries, which were identified following Jun's (Jun 2005a) revised proposal. A canonical prosodification of an example tongue twister is provided in (5.1). In this example, each lexical word is divided by a space.

(5.1) [[*jeogi issneun*]_{AP} [*malttug-i*]_{AP} [*mal mael*]_{AP} [*malttug-inya*]_{AP}]_{IP} [[*mal mot mael*]_{AP} [*malttug-inya*]_{AP}]_{IP}

Is the stack there the one to chain a horse or not?

5.3. Results and discussion

5.3.1. Speech errors

A total of 1,301 speech errors were identified in 888 sentences embedded in 863 repetitions of different tongue twisters. Of the 1,301 speech errors, 541 were categorized as errors of anticipation, 374 as errors of perseveration, 211 as errors of anticipation and perseveration, and 179 as deletion, insertion, or lexical substitution. Thus, the majority of errors were contextual errors (86.24%). The error rate, calculated as number of tongue twisters with errors out of the total number of tongue twisters produced, ranged from

15.38% to 44.23% across the 20 participants. The average error rate was 32.76% ($SD = 8.42\%$).

Of the 1,126 contextual errors identified, 51 (4.53%) involved whole syllable movement, 247 (21.94%) involved vowels, but the majority (73.53%, $N = 828$) involved consonants. The proportion of errors involving vowels was high in comparison to English, where only 5.57% of the errors elicited involved vowels (Chapter III). Also unlike English, the 828 consonantal errors occurred more often in syllable coda position ($N = 486$) than in syllable onset position ($N = 342$). Most of the contextual errors had their source in an adjacent syllable (77.89%), possibly an artifact of the definition of a source as the nearest possible candidate. Other contextual errors clearly had their source in a nonadjacent syllable (21.31%). A few errors appeared to have their source within the same syllable, representing an anticipation or perseveration of the consonantal coda ($N = 2$) or onset ($N = 7$).

5.3.2. Supralexical prosodic units

The 757 sentences with contextual errors were embedded in a total of 714 repetitions of different tongue twisters. These tongue twisters were then prosodified. Table 5.1 shows the number and size of the different supralexical prosodic units that contained one or more contextual errors.

Table 5.1. Number and length of the 3 different supralexical prosodic units that contained one or more contextual errors.

Supralexical prosodic units	<i>N</i>	Range	Average (<i>SD</i>)
Accentual phrases (AP)	911	2 – 6	3.80 (0.97)
Intermediate phrases (ip)	113	3 – 22	8.86 (4.08)
Intonational phrases (IP)	858	4 – 32	2.95 (5.31)

Note. The number of syllables is the measure of phrase length.

The majority of the prosodified tongue twister productions had more than one IP, but 34.35% were uttered under a single IP. Only 18.10% of IPs had internal ip boundaries. Recall that the ip unit emerges from focus marking or with complex syntax so Korean sentences are not exhaustively prosodified into IP, ip, and AP. Rather, there are always many fewer ips than IPs or APs. That said, the finding that 18.10% of IPs with errors contained ips with errors is interesting because it suggests that speakers likely made frequent use of focus marking, perhaps in an attempt to make the tongue twister repetition task easier. Unlike with ips, IPs can be exhaustively prosodified into constituent APs. The table shows that 858 IPs contained at least one contextual errors; 30.06% contained more than one error. The table shows that 911 APs contained at least one contextual error; 14.24% contained more than one contextual error.

Table 5.1 also indicates that the average size of APs with errors was quite similar to the reported size of these units in Jun (2005b), who notes that most APs are 3-4 syllables long. Jun also reported that the number of syllables in a Korean IP typically ranges from 12 to 15 syllables. The average length of IPs with errors in the present experiment was 12.95, which is within this range. Thus, the comparison with Jun's

figures suggests that errors did not influence participants' prosodification of the tongue twisters.

5.3.3. Distribution of errors in supralexicalexical prosodic units

As in Experiments 1 and 2 (Chapters III and IV), scatterplots in Figure 5.1 were used to evaluate the distribution of errors across positions in the supralexicalexical prosodic units. Error position was defined with respect to the position of the syllable within which the error occurred, counting from the left-edge of the unit to the right. Error position was calculated independently for each error within a unit. Unit length was calculated as number of syllables. This coding of position and length was meant to allow for more direct comparisons to English: most of the English words featured in the tongue twisters were monosyllabic. Also, the notion of word is a bit more problematic in Korean than in English because Korean is an agglutinative language.

As before, regression lines were drawn for each scatterplot. Recall that the slope of the error-position by syllable-number regression line represents the overall distributional pattern of speech errors within a phrase such that when the slope is larger than 0.5 more errors have occurred towards the end of a phrase, and when the slope is smaller than 0.5 more errors have occurred towards the beginning of a phrase.

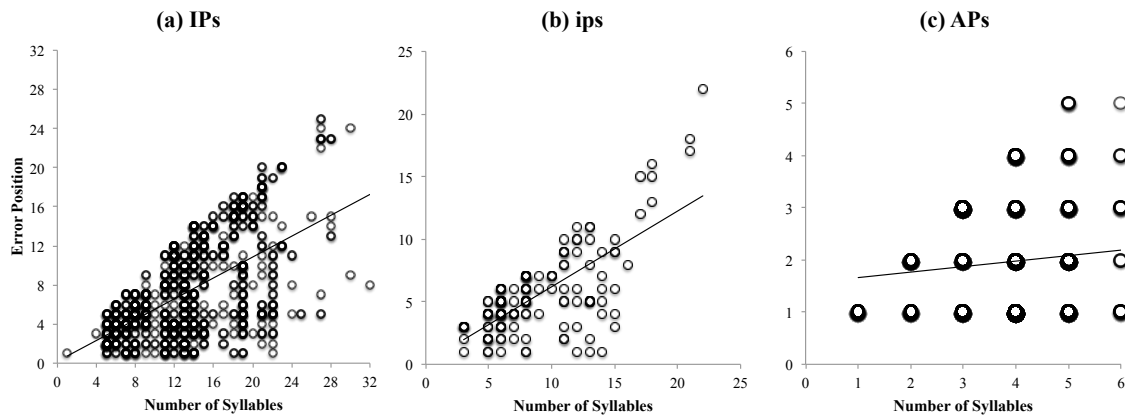


Figure 5.1. Position of syllables containing speech errors (error position) as a function of the number of syllables in the IP, ip, and AP. The circles represent errors. Darker circles correspond to greater numbers of errors than lighter circles. The black lines represent the error-position by syllable-number regression lines.

The slope of the error-position by syllable-number regression line for IPs was 0.53 ($R^2 = 0.35$), and that for ips was 0.61 ($R^2 = 0.50$). These slopes indicate that slightly more contextual errors occurred towards the end of a unit than toward its beginning. The values also clearly suggest that errors were more evenly distributed within Korean IPs and ips than within the English IP. In fact, the slope associated with errors in the Korean IP was very close to the slope we would expect if the errors were randomly distributed across the unit ($= 0.5$). Although this may indicate that Korean ip/IPs function differently from English IPs in the planning process, the discrepancy between the languages could also be due to the different ways in which position and length were coded—in syllables for Korean, but in words for English.

The distribution of errors within the Korean APs was completely distinct from that observed for the larger supralexical prosodic units. As Figure 5.1c shows, the slope of the error-position by syllable-number regression line in APs was 0.11 ($R^2 = 0.01$). This

nearly flat slope indicates that more errors occurred towards the beginning of the unit than towards its end.

To better compare the distribution of errors within the smaller and larger supralexic prosodic units, errors were summed across the first and second halves of each AP¹⁴, and then presented as a function of position (initial, medial, final) within the larger ip/IP unit. AP position within an ip was coded when an error occurred in an utterance with a noted ip boundary. Otherwise, AP position was coded with respect to the IP boundary. The results are shown in Figure 5.2.

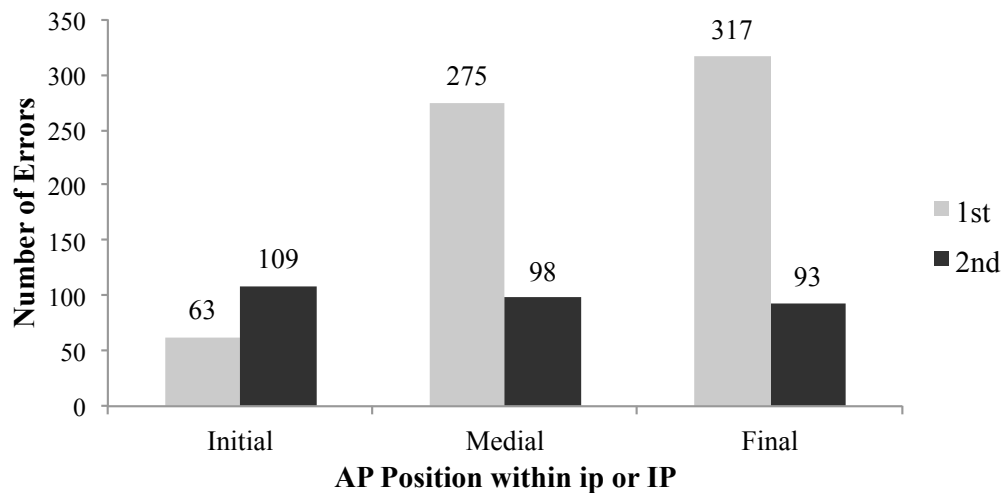


Figure 5.2. Number of errors as a function of error position in the AP (1st half, 2nd half) and position within the ip/IP (Initial, Medial, or Final)¹⁵.

¹⁴ Only 1,067 contextual errors (94.76%) were included in this analysis. Fifty-nine of them were excluded because these errors occurred in APs with one syllable or in APs whose boundaries corresponded with either ip or IP boundaries.

¹⁵ The figure excludes errors that occurred in the middle of APs with odd numbers of syllables since it was unclear whether to consider this position as belonging to the first or second half of the AP. To ensure that the exclusion of these APs did not introduce a bias into the ip/IP positional results shown in Table 5.2, the distribution of medial errors as a function of position within the larger unit was independently assessed for in these units. Out of 112 errors in APs with odd numbers of syllables, 16 were in ip/IP-initial APs, 23 were in ip/IP-medial APs, and 73 were in ip/IP-final APs.

Another pattern that is evident in Figure 5.2 is that the distributional pattern of errors within an AP varied as a function of its position within the larger supralexical prosodic unit. In particular, fewer errors occurred in AP-initial position compared to AP-final position when these APs occurred in ip/IP-initial position. The unique error distributions across ip/IP position may result from the influence of the ip/IP on the phonological encoding within APs. This point will be discussed in more detail below.

The AP position within an ip or IP also affected the overall number of contextual errors in APs. Specifically, 17.62% of the contextual errors occurred in ip/IP-initial APs, 37.11% in ip/IP-medial APs, and 45.27% in ip/IP-final APs. Thus, many more errors occurred in ip/IP-final APs than in ip/IP-medial APs, even though ip/IP-medial position usually included more than one AP. This pattern is similar to the cumulative error pattern observed in the English IP. Note, though, that the cumulative error pattern only really emerges in the Korean data when one identifies the appropriate unit of analysis, which seems to be the AP.

The next analysis followed up on the preceding one. The goal was to further investigate the distribution of different types of contextual errors within the ip/IPs as a function of sentence position to ensure that the relevant unit was prosodic (ip or IP) rather than sentence-based. First, the data were divided into ip/IP with final boundaries that coincided with the sentence final boundary (sentence-final ip/IPs), and those where the final boundary corresponded to the internal boundary of the following ip/IP, but not to the sentence boundary (nonfinal ip/IPs). The total number of errors was then calculated

for each fourth of the IP, whose length was defined in number of syllables. The results are shown in Figure 5.3¹⁶.

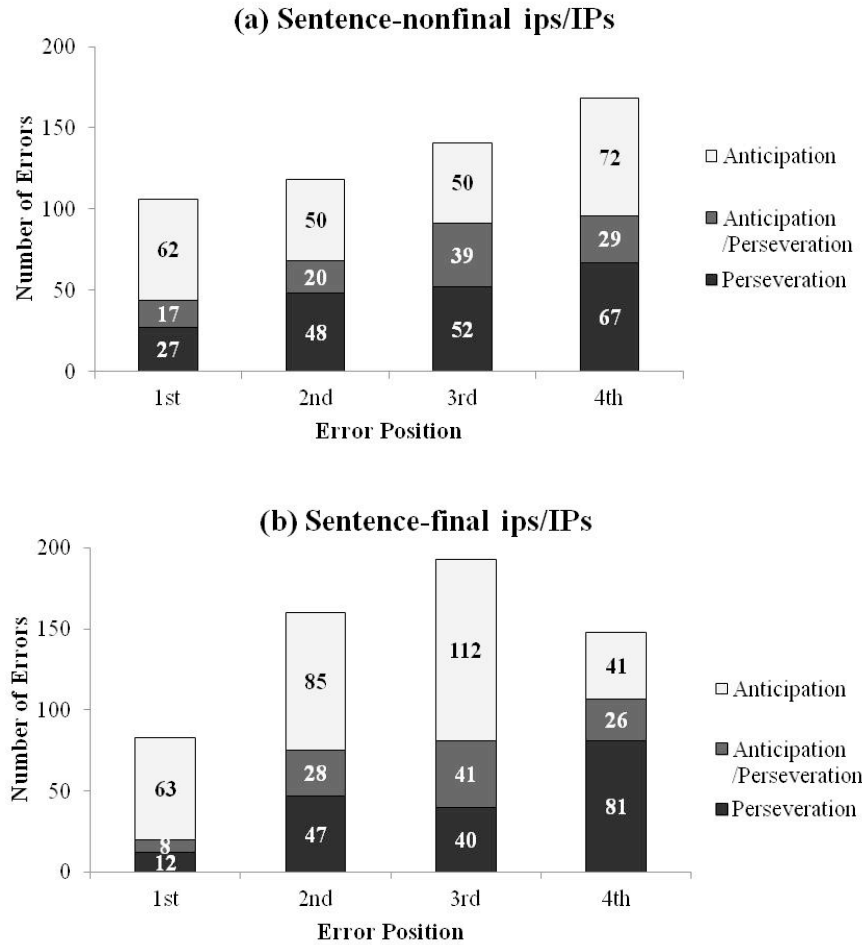


Figure 5.3. Number of anticipatory and/or perseveratory errors in ip/IPs as a function of error position

Figure 5.3 confirms that the cumulative error pattern exists independently of ip/IP position within the sentence. Figure 5.3 also shows that, like in English, the error distribution within ip/IPs varies as a function of their position in a sentence. In particular, the final position effect was observed here as in English, where fewer errors occur at the

¹⁶ Nine contextual errors were excluded for further analyses since these errors occurred in ips or IPs with less than 4 syllables.

end of the unit when the unit is also in sentence-final position. The specific break down by error types shown in the Figure indicates that, as in English, the final effect is due to a drop off in the number of anticipatory errors that occur in sentence-final position. Note that, also like in English, a large proportion of errors that occur in the last quarter of sentence-nonfinal ip/IPs are anticipatory. This suggests interference from preparing elements in the following ip/IP for execution.

5.4. General discussion

In summary, the error patterns from the current experiment showed similarities and differences with those from adult English (Chapter III). The primary differences between the languages were in the overall characteristics of the errors. First, the average error rate in the Korean data was 28.46%, which was much higher than the average error rate of 8.1% in the English data. There was also a noticeable difference in the distribution of contextual errors across syllable positions. Specifically, the Korean data showed a much higher proportion of errors that involved the syllabic nucleus and, especially, the syllabic coda than the English data. By contrast, the distribution of contextual errors within the highest-level supralexical prosodic units in Korean was similar to the pattern observed for the comparable units in English. Specifically, both the cumulative error pattern and a final position effect were observed. An additional, very interesting finding from the current experiment was the distribution of errors within APs, a unit that has no obvious equivalent in English. Overall, more errors occurred in the first half of an AP than in the second half, though this pattern was not maintained when APs occurred at the beginning of ip/IPs. The implications of all of these findings are discussed below.

5.4.1. Overall patterns of speech errors in Korean

The finding that the average error rate across participants was higher in Korean than in English was likely an artifact of experimental design. In particular, all the stimuli in the current experiment were tongue twisters; no filler non-tongue twisters were used. The constant repetition of tongue twisters could have tired the Korean-speaking participants more than adult English-speaking participants, who produced filler sentences in between repetitions of tongue twisters. In addition, most Korean tongue twisters are much longer than English tongue twisters, and they are certainly much longer than the stimuli used in other error-eliciting experiments (e.g., Dell (1986) used 2-word tongue twisters; Croot et al. (2010) used 4-word tongue twisters). The longer tongue twisters may also have contributed to the higher error rates in Korean as compared to English.

Whereas the higher average error rates in Korean may have been due to the task, the finding that Korean contextual errors more often involve the syllabic nucleus or coda than the syllabic onset may be due to language structure. The common Korean tongue twisters used in the current experiment rely on alternations in syllabic nuclei (e.g., *nae.gea.geu.lin gu.leum geu.lim...*), rhymes (e.g., *bag.beom.bog.gun-eun bam beot.kkot.no.li.leul...*), and codas (e.g., ... *gawg.jin.gwang gwan.gwang.gwa.jang*) to confuse speakers. In contrast, the common English tongue twisters used in Experiment 1 rely on alternations in syllable onsets (e.g., *big black bear bleed blood*) to confuse speakers. The structural differences in tongue twisters in the two languages may have something to do with their different rhythm structures. Recall that Korean is syllable-timed whereas English is stress-timed. In a syllable-timed language each vocalic nucleus

must be fully realized, but in a stress-timed language full vowels typically alternate with reduced vowels. Thus, rhyme-based repetitions promote confusion of adjacent elements in a syllable-timed language (assuming structure preservation), but would require confusions of non-adjacent elements in a stress-timed language. For this reason, it may be more possible to build rhyme confusing tongue twisters in Korean than in English, leading to the language-specific differences then observed in what elements within the syllable are more error prone.

It may also be that the language rhythm structure affects what unit is fundamental to the planning process. Whereas in English it is assumed that phonemes are serially ordered, it could be that syllables are the units that are serially ordered in a syllable-timed language. This is the argument put forth by Chen and colleagues (1999, 2000; Chen, Chen & Dell, 2002; O'Seaghdha & Chen, 2009; O'Seaghdha, Chen, & Chen, 2010), based on the higher rate of syllable errors to contextual errors in spontaneously produced Mandarin Chinese compared to English. Studies on speech segmentation provide additional support for the idea. In syllable-timed languages of French and Mandarin Chinese, the evidence suggests that syllable boundaries are clear and provide cues for speech segmentation (e.g., Mehler, Dommergues, Frauengelder, & Segui, 1981; Tseng & Tsao, 1994; Tseng, Huang, & Jeng, 1996; Tseng, 1997). In stress-timed English, lexical stress provides the fundamental cue to word boundary location (e.g., Cutler, Mehler, Norris, & Segui, 1989). If syllables rather than phonemes are retrieved and serially ordered during the phonological encoding process, the chance for syllables to be misselected will increase. This was the finding in the present experiment. Of course, the stimuli in the present experiment were not balanced for number of confusable phoneme

pairs across syllable positions, and so the idea that syllables may be more important than phonemes in Korean speech planning remains to be tested.

5.4.2. Korean supralexic prosodic units in speech planning and production

A central finding of the current experiment was that the distributional pattern of contextual errors differed for APs and ip/IPs. Specifically, errors were more likely to occur in initial or early position in APs than in late position, but errors were more likely to occur towards the end of ip/IPs than at the beginning. These results are interpreted to suggest that Korean APs and ip/IPs have different roles in speech planning and production. In particular, we conclude that IPs in Korean function similarly to those in English—as the largest domain within which speech planning occurs—and that APs function more like prosodic or lexical words in the encoding process.

APs are nested within IPs in Korean, just as phonological words are nested within IPs in English. Interestingly, the higher error rates in AP-initial syllables are also analogous to the (prosodic and/or lexical) word-level distributional pattern in English wherein higher error rates are observed in word-initial position than in any other positions within a word. If we assume that the so-called word-initial effect in English is due to their unique structural status of word onsets (see e.g., Shattuck-Hufnagel, 1992), then we might explain the AP-initial effect in a similar manner. Recall that an AP has only one of two tonal patterns, LHLH or HHLH, and the pattern that is chosen is determined by the AP-initial segment (i.e., H for aspirated, tensed, /s/, or /h/, and L for others). This feature of the AP might have consequences for the planning process. For example, the more predictable aspects of the IP tonal structure might be established first

and the associated syllables ordered within the tonal frames before AP-initial syllables are assigned. If we assume that AP-initial elements are slotted into the AP frames at a later stage in the encoding process, and that later stages in the process are more error prone than earlier stages (due to the temporal decay in overall IP activation), then this could be why a disproportionate number of errors occur in AP-initial position, particularly when these occur late in the IP.

An alternative explanation for the AP-initial effect is inspired by Dell and colleagues' (1993) PDP model. Recall that in this model, execution is based on two types of activation: one that arises from the phonologically encoded lexical representations and one due to contextual activation, specifically activation from preceding segments in the phonetic plan. The word-initial effect then is due to the lesser activation associated with word-onset segments in this plan. If we assume that the AP is the domain within which the final stages of speech planning occur and that contextual activation plays a role in maintaining serial order during execution, then initial AP position will be similarly vulnerable to errors since it does not benefit from contextual activation.

The above explanations assume that the AP is relevant to planning either because it serves as a tonal frame within which syllables are slotted during phonological encoding or because it serves as the domain of planning for execution. However, an alternative explanation for the AP-initial effect does not need to assume the AP as an independent structure relevant to planning. Instead, this effect could be explained in terms of lexical representations, that is, exactly in the same way as the word-initial effect is explained in English. This last possibility is due to the fact that although the AP is defined as a supralexical prosodic unit because it can span more than one lexical word, more often

than not the unit corresponds with a single lexical word (Jun, 2005b). Thus, before finalizing our conclusions regarding the status of different supralexical prosodic units in speech planning, it is necessary to establish whether the error distribution observed in the present experiment is in fact attributable to the AP, a prosodic unit, rather than to lexical representations.

CHAPTER VI

EXPERIMENT 4: SPEECH ERRORS
IN NONSENSE TONGUE TWISTERS
PRODUCED BY KOREAN SPEAKERS

6.1. Introduction

Experiment 3 examined speech errors in Korean and found that more contextual errors occurred in the first half of an AP than in the second half and that more than half of these early errors were in AP-initial position. This AP-initial effect was interpreted to mean that the Korean AP functions similarly to syllable structure in English words by providing frames for serial ordering within the larger IP unit. However, since the Korean AP usually extends over just one content word, it was not possible to determine with confidence whether the distribution of errors within the AP should be attributed to tonal frames or to the structure inherent in lexical representations. The current experiment was conducted to provide a stronger test of the hypothesis that the prosodic unit is the relevant unit of speech planning in Korean.

In the previous chapter, two different explanations for the AP-initial effect were provided, both of which assumed the psychological reality of the prosodic unit in speech planning and production. These explanations extended different model accounts of the word-initial effect in English to the AP-initial effect in Korean. The different models in turn represented two different views on the nature and role of words in speech planning and production. Shattuck-Hufnagel's (1987, 1992) frame-based model, from which our

tonal structure explanation was derived, assumes that speech planning occurs within larger units (e.g., phrases or utterances), but that the serial ordering of phonemes is largely dependent on syllable frames that are embedded in metrical structures (prosodic words), defined by lexical stress patterns. Dell and his colleagues' (1993) PDP model, from which our contextual activation explanation was derived, argues that phonological planning need not reference word-internal structure. According to this model, serial ordering is stored as part of the lexical representation and is implemented based on activation from higher-level units. Once the appropriate phonological string has been activated, phonetic spell-out can proceed from left-to-right within the domain defined by the length of the lexical item and strengthened by contextual activation, such that subsequent elements receive additional activation from already spelled-out elements in the word. The two planning processes can be extended to Korean APs as illustrated in Figure 6.1 and described below.

With respect to the tonal structure explanation for an AP-initial effect, let us assume that speakers plan their speech within the larger domain of an IP so that all candidate syllables¹⁷ within the IP are retrieved and stored in a short-term buffer for later position assignment. Let us further assume that when the candidate syllables are stored in the buffer, AP boundaries are marked and each syllable is tagged by its position within an AP. During phonological encoding, the tagged elements are slotted into the appropriate frame from left-to-right. The elements tagged for initial position are however held in reserve, as the identity of these elements will change the initial tonal specification of the AP (see Chapter V). As activation of the IP decays with time, so too does the activation

¹⁷ One of the underlying assumptions in this chapter is that syllables are the units that Korean speakers retrieve and serially order from the lexicon as in other syllable-timed languages such as Mandarin Chinese (see Chapter V).

of each item not yet spelled out for production. This results in the AP-initial effect and the increasing strength of this effect across the IP.

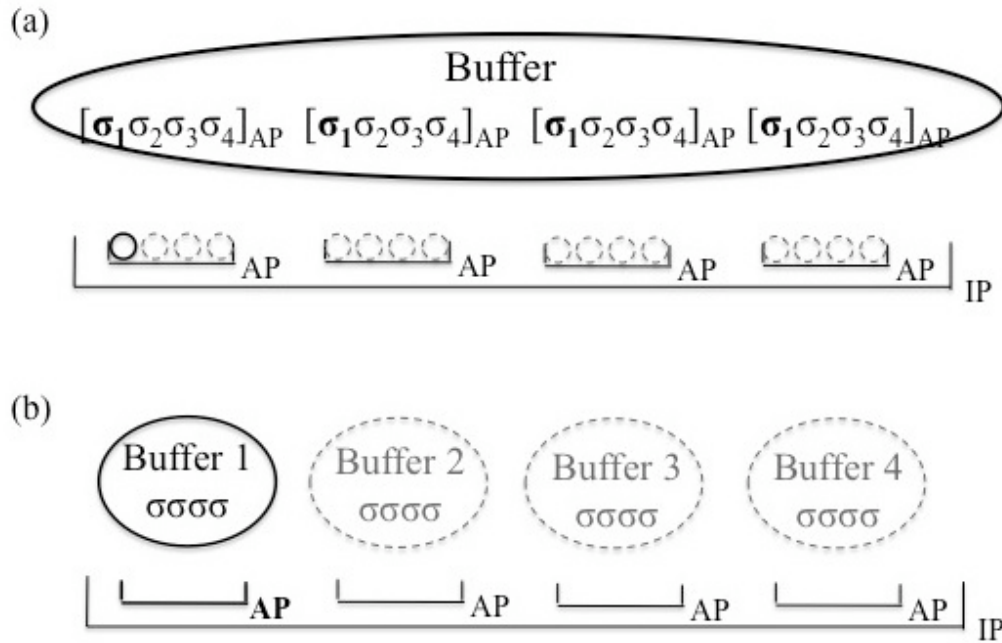


Figure 6.1. Two possible roles for the AP in speech planning. The top panel (a) represents the AP as prosodic frame within which items are serially ordered according to their tags. This possibility follows from the assumptions of Shattuck-Hufnagel’s (1987, 1992) extended scan-and-copy model. The bottom panel (b) represents the AP as the phonetic spell-out domain with the unit’s phonological material activated by a higher-level representation. This possibility follows from the assumptions of Dell et al.’s (1993) PDP model. The solid and darker lines and letters indicate the items that are currently activated. Further details are in the text.

With respect to the spreading activation account of the AP-initial effect, let us continue to assume that speakers plan their speech within the larger IP domain, but that spell-out processes take place more locally within the domain of the AP. Thus, unlike in the tonal structure explanation, only syllables within a single AP are activated during the final stages of speech planning. It is further assumed that as syllables within each AP are phonetically spelled out (e.g., articulatory goals are specified and motor routines

retrieved), they activate the subsequent syllable within the local domain. Since contextual activation is only cumulative within one AP, activation levels are lower for the initial elements within the local domain and so are more error prone.

Note that the different accounts for the AP-initial effect also make different predictions regarding structure preservation. Structure preservation is more naturally a feature of the tonal structure account since the assumption is that all elements in the buffer are tagged for position within the AP. By contrast, in the spreading activation account, the only elements activated for encoding at any given moment in time are those in the AP domain. Thus, contextual errors are more likely to be affected by another syllable in the same domain than by a syllable from another AP that shares position information with the activated AP.

To summarize, the current experiment was designed to provide a strong test of the hypothesis that the AP is a planning domain in Korean. Speakers were given practice with two-word APs to ensure that the AP was treated as a supralexical prosodic unit. Then non-words were used in an error elicitation task to control for influences from lexical representations. The non-words participated in tongue twisters that were developed to manipulate the location of errors with respect to the AP boundary. If APs are relevant to the planning process, then information about the location of the source and error will be important for determining whether the AP functions as a planning frame as in Figure 6.1 (a) or is merely the local domain within which the last stages of planning proceed as in Figure 6.1 (b).

6.2. Methods

6.2.1. Participants

Twenty native speakers of Korean participated in the current study (5 males and 15 females). All participants spoke the Kyoung-gi dialect of Korean, which is the standard dialect of Korean. The participants were adult native Korean speakers who lived in Eugene, Oregon at the time of the experiment, and their age ranged from 20 to 46 ($M = 28.9$, $SD = 8.2$). No participants had lived outside of Korea for more than 5 years. Three of the 20 participants had also participated in Experiment 3 (Chapter V).

6.2.2. Stimuli

In order to minimize the influence of lexical representations on the distribution of contextual errors within the AP, nonsense tongue twisters were created. Each tongue twister was constructed out of syllable strings (hereafter SS) and tagged with grammatical particles to render the syntactic structure shown in (6.1). The purpose of using pseudo-grammatical tongue twisters was to make speakers produce sentence-like prosody rather than list-like prosody. Practice was used to ensure that all tongue twisters would be produced as 5 APs embedded in a single IP (see 6.2.3). The particular sentence structure shown in 6.1 was used because all of the particles in this structure consist of vowels only. These were used so that error-prone consonants in the SSs would be less affected by the meaningful morphemes.

- (6.1) *SS₁-ui* *SS₂-i* *SS₃-ui* *SS₄-e* *noh-yeo-issda*¹⁸
 SS₁-gen *SS₂-nom* *SS₃-gen* *SS₄-loc* place-passive-be.int
 ‘SS₁’s SS₂ is on SS₃’s SS₄’

To investigate whether the AP provides a tonal frame for speech planning within the IP or is simply the domain of phonetic spell-out, the position of error-prone syllables relative to AP boundaries was manipulated. For half of the stimulus tongue twisters (Type 1 and 3) a pair of error-prone syllables straddled an AP boundary, and for the other half (Type 2 and 4) a pair of error-prone syllables occurred within the domain of a single AP. Two other factors were controlled: the distance between error-prone syllables, and the total number of syllables in a single tongue twister sentence. The manipulation of these three factors yielded four types of tongue twisters as shown in Table 6.1.

Table 6.1. Four types of tongue twisters and the precise manipulation of three factors.

Type	Syllable-position	Distance	Length	SSs (SS ₁ / SS ₂ / SS ₃ / SS ₄)
1	Across AP	4	Long	<u>σ₁₁</u> σ σ σ / <u>σ₁₂</u> σ σ σ / <u>σ₂₁</u> σ σ σ / <u>σ₂₂</u> σ σ σ
2	Within AP	2	Long	<u>σ₁₁</u> σ σ <u>σ₁₂</u> / σ σ σ σ / <u>σ₂₁</u> σ σ <u>σ₂₂</u> / σ σ σ σ
3	Across AP	2	Short	<u>σ₁₁</u> σ / <u>σ₁₂</u> σ σ σ / <u>σ₂₁</u> σ / <u>σ₂₂</u> σ σ σ
4	Within AP	2	Short	<u>σ₁₁</u> σ σ <u>σ₁₂</u> / σ σ / <u>σ₂₁</u> σ σ <u>σ₂₂</u> / σ σ

Note. A distance represents the number of syllables between two error-prone syllables. As to a sentence length, long sentences have 24 syllables, and short sentences have 20. Syllables with error-prone consonants are underlined and in bold. The position of each syllable is marked with subscripts like **σ_{ab}**, the details of which manipulation will be mentioned in the text.

¹⁸ All Korean letters here and in Appendix B for Experiment 3 were romanized by Notification No. 2000-8 of the Ministry of Culture and Tourism. The information is found in the website of the National Institute of the Korean Language (http://www.korean.go.kr/eng_new/document/roman/roman_01.jsp).

All syllables in the target SSs had a CVC structure, where V was always the low vowel /a/. Each tongue twister had 4 error-prone syllable-initial consonants. Syllable rhymes were held constant (/al/). The consonant pairs /t^h-/t̥/, /p/-/p^h/, /t͡ʃ/-/s/, and /k^h-/k̥/ were used to create confusion through alternation. These pairs were chosen as they were the most error-prone pairings in Experiment 3. In Table 6.1, each error-prone syllable is presented with two subscripts as σ_{ab} : the first subscript represents the order of pair, and the second represents the order of consonant within a pair. Let us calculate the number of cases in which 4 different consonants occur in 4 different positions¹⁹. First, 2 out of 4 pairs need to be chosen and ordered for each tongue twister, so this yields 12 different cases (4 possible pairs for the first pair (σ_{1x}) X 3 possible pairs for the second pair (σ_{2x})). Since 2 consonants in each pair can be placed in 2 different positions (σ_{x1} & σ_{x2}), this yields 4 different cases (2 for ordering consonants in pair 1 \times 2 for ordering consonants in pair 2). Therefore, a total of 48 different combinations per type are created. Each participant produced a total of 384 tongue twisters in one of the two pre-determined randomized order: 48 different combinations of error-prone consonants \times 4 types \times 2 repetitions.

Three filler syllables (/han/, /maŋ/, /nam/) were used in non-error-prone positions for speakers to make sufficient length of APs (σ without subscripts in Table 6.1) without making any meaningful words. The example tongue twisters are presented in (6.2) below. Note that /t̥/, /ŋ/, /t͡ʃ/, and /t̄/ in syllable coda position are represented here as *tt*, *ng*, *ch*, and *ss*, respectively, in accordance with the romanization conventions put forth by the Korean Ministry of Culture and Tourism.

¹⁹ Two consonants of a single pair of consonant were always placed next to each other.

(6.2)

- a. Type 1: tal.han.mang.nam.ui ttal.han.mang.nam.i
chal.han.mang.nam.ui sal.han.mang.nam.e noh.yeo.iss.da
- b. Type 2: tal.han.mang.ttal.ui nam.han.mang.nam.i
chal.han.mang.sal.ui nam.han.mang.nam.e noh.yeo.iss.da
- c. Type 3: tal.han.ui ttal.han.mang.nam.i chal.han.ui
sal.han.mang.nam.e noh.yeo.iss.da
- d. Type 4: tal.han.mang.ttal.ui han.nam.i chal.han.mang.sal.ui
han.nam.e noh.yeo.iss.da

6.2.3. Procedure

The elicitation procedure in Experiment 4 was largely identical to that used in Experiment 3 (Chapter V). The tongue twister repetition task took just over an hour to complete. Once again, participants performed the task in a quiet laboratory room and the entire session was digitally recorded using a Marantz PMD660 for later analysis using Shure ULXS4 standard wireless receiver and lavalier microphone. However, unlike in the previous experiment, each participant engaged in 10 minutes of practice before attempting to produce the nonsense tongue twisters in the elicitation task. The practice familiarized participants to the sentence structure, and ensured that they would be able to produce the nonsense tongue twisters with the intended prosodic structure (five APs embedded in one IP).

The materials for the practice session were all meaningful sentences that had sentence structures identical to the stimulus tongue twisters (see (6.1)). The meaningful sentences varied in length from 3-syllable APs to 5-syllable APs. Moreover, 10 of the 40

APs (4 different APs \times 10 sentences) included in the practice sentences contained 2 words. This design feature was meant to encourage participants to abstract supralexical prosodic units instead of simply equating the AP with a content word. During practice, participants were overtly instructed to produce each sentence without any noticeable pause. The goal of this instruction was to ensure that the sentence be produced under a single IP. Participants were also not allowed to use any emphasis or contrastive focus. The goal of this instruction was to prevent sentences from being produced with intervening ip boundaries. The participants repeated the practice materials up to three times until the experimenter was satisfied with their phrasing.

After the practice session, participants were informed that all target sentences would be nonsense tongue twisters that had the same sentence structures as the practice materials. They were further asked to use the same phrasing as in the practice session. The author, a native Korean speaker, monitored participants' productions during the tongue twister elicitation task and asked participants to any sentences that were not produced with the desired prosodic phrasing. Fortunately, this instruction did not need to be given very often once participants became familiar with the nonsense syllable strings.

6.2.4. Coding

The author noted speech errors *in situ*, and then listened to all 7,680 recorded sentences (384 sentences \times 20 speakers) to confirm error occurrence. Errors were simply categorized as contextual or non-contextual. Unlike in the other experiments that used real-word tongue twisters, the current experiment did not code for errors of anticipation and perseveration because the principle question was whether contextual errors were

more likely to occur across an AP boundary or within an AP. In addition to error coding, 10% of the sentences produced in the tongue twister repetition task were randomly selected for prosodic transcription. The goal was to check that all speakers had in fact prosodified the sentences in the manner specified. This spot check revealed that all participants followed the phrasing that had been practiced such that each sentence was produced with 5 APs embedded within a single IP.

6.3. Results and discussion

6.3.1. General characteristics of speech errors

Overall speech errors occurred in 15.7% of the stimulus sentences. Individual error rates ranged from 7.03% to 29.17% ($SD = 7.48\%$). A total of 1,206 speech errors were identified. The majority of these were contextual errors (96.35%). In addition, the majority of contextual errors (67.74%, $N = 1,206$) occurred in the error-prone syllables. These will be referred to as target syllable errors, and they are the focus of the analyses presented below.

6.3.2. Effect of error-prone syllable position on error occurrence

Recall that the prediction was that more errors would occur when error prone syllables straddled an AP boundary if the AP provides a prosodic (tonal) frame for phonological planning. By contrast, if the AP serves as a local domain for the final stages of speech planning, then the prediction was that more errors would occur when error prone syllables were in the same AP. These predictions were tested in Poisson regression models with error-prone syllable position (across-AP vs. within-AP) as a predicting

variable. The dependent variable was the mean number of target syllable errors calculated for each participant across stimulus sentences and repetitions. The result showed that participants made significantly more target syllable errors when error-prone syllables straddled an AP boundary ($M = 22.45$, $SD = 13.78$) than when these occurred within the domain of an AP ($M = 18.40$, $SD = 11.75$), Wald $\chi^2 = 8.00$, $df = 1$, $p = .005$, $B = 0.20$.

To confirm that the effect of error-prone syllable position remained when the distance between error-prone syllables was controlled, the analysis was rerun with the mean number of target syllable errors elicited only during repetitions of Type 3 and Type 4 stimuli. Again, the analysis confirmed that participants made significantly more target syllable errors when error-prone syllables straddled an AP boundary ($M = 13.60$, $SD = 6.35$) than when these occurred within the domain of an AP ($M = 9.25$, $SD = 5.65$), Wald $\chi^2 = 16.36$, $df = 1$, $p < .001$, $B = 0.39$. These results are shown in Figure 6.2.

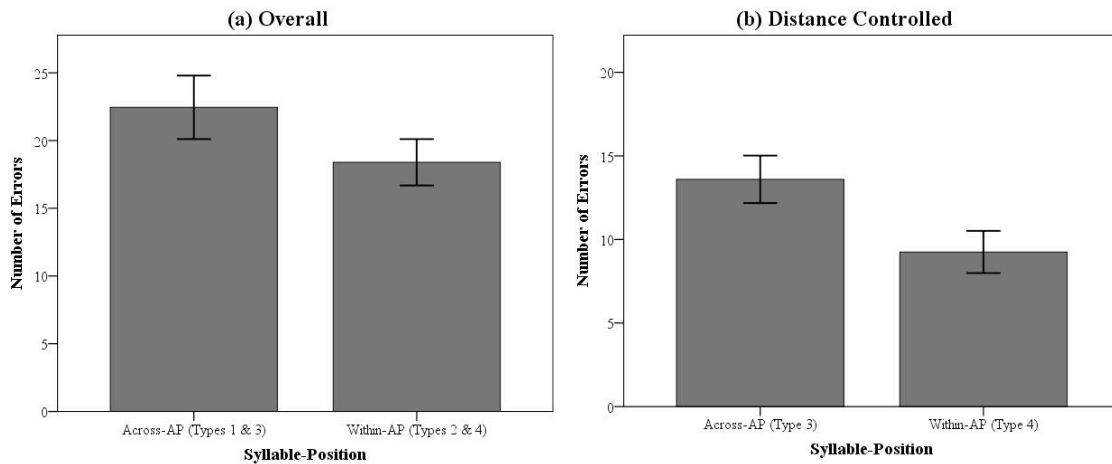


Figure 6.2. The mean number of target syllable errors as a function of syllable position (Across-AP vs. Within-AP). Panel (a) shows the mean number of errors across tongue twister types, while (b) shows the mean number of errors in Types 1 and 3 where distance between the source and error were controlled. Each error bar represents ± 1 standard error.

In sum, the results indicate that significantly more errors occurred when error-prone syllables straddled an AP boundary than when these occurred within the domain of an AP even when the distance between the source and error was controlled. These results support the hypothesis that the Korean AP provides a prosodic frame for planning within the larger IP domain. Specifically, the error and source more often shared position information across APs than were confused within an AP. Next, the distribution of errors within the IP is examined to test for the expected cumulative error pattern that provides evidence of the IP as a planning domain.

6.3.3. The distribution of target syllable errors within the IP

To investigate the distribution of errors within the IP, the number of errors within each AP was summed, and then plotted as a function its position within the IP. Note, however, that the potential location of errors varied as a function of the syllable-position factor. When the error-prone syllables straddled an AP boundary (Types 1 & 3), errors could occur in any one of the first 4 APs²⁰. However, when the error-prone syllables occurred within a single AP (Types 2 & 4), then target syllable errors could only occur in the first and third AP. For this reason, the effect of IP position on number of errors was investigated separately by the syllable-position factor (across-AP vs. within-AP).

Poisson regressions were conducted to test whether the number of target syllable errors varied as a function of AP-position within the IP. The predicting variable, AP-position included 4 levels (1st, 2nd, 3rd, and 4th position) for errors elicited using tongue twister Types 1 and 3, and 2 levels for errors elicited using tongue twister Types 2 and 4

²⁰ Because the last AP was a real word (*noh.yeo.iss.da*), identical across all tongue twisters, and used to render a pseudo-grammatical structure, no target syllable errors could occur in this position. The 5th AP was therefore excluded from the analyses.

(1st and 3rd position). The dependent variable, number of target syllable errors, was calculated for each participant across stimulus Types and repetitions within the factor of interest. The results indicated that participants made more errors in later APs for both Types of stimuli, Wald $\chi^2 = 161.10$, $df = 3$, $p < .001$ for Types 1 and 3, and Wald $\chi^2 = 49.14$, $df = 1$, $p < .001$, $B = -0.75$ for Types 2 and 4. In particular, the unstandardized coefficient values of 1st, 2nd, and 3rd APs for Types 1 and 3 confirmed that participants made significantly more errors in the 4th AP compared to other APs in the IP (B values of 1st, 2nd, and 3rd APs were -1.36, -0.60, and -0.50, respectively). The mean number of errors is shown as a function of AP position with the IP in Figure 6.3.

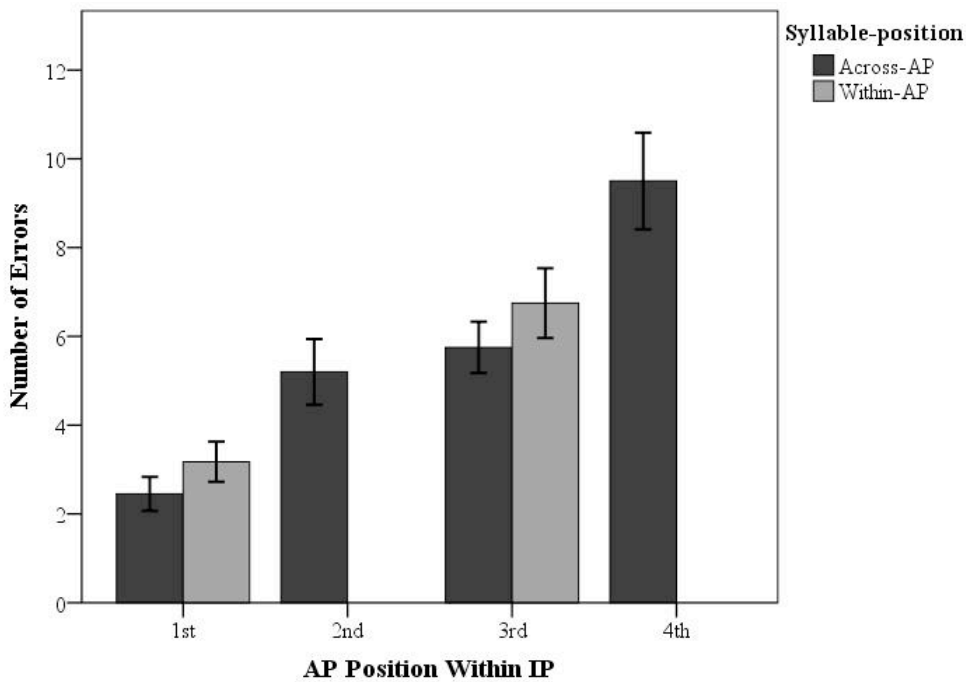


Figure 6.3. The mean number of target syllable errors for each AP position within the IP. Note that target syllable errors could occur in any of the 4 APs for Type 1 & 3 (Across-AP) tongue twisters, but that target syllable errors could only occur in the 1st and 3rd AP for Type 2 & 4 (Within-AP) tongue twisters. Each error bar represents ± 1 standard error.

Figure 6.3 shows that fewer errors occurred in the first AP than in later APs, and that the pattern was gradient and highly reminiscent of the cumulative error pattern that has been observed in each of the experiments presented in the current dissertation.

Altogether, the results of the analyses indicated that AP position within the IP had the expected effect on the distribution of contextual errors. The results demonstrated once again that errors are cumulative across the AP, even when the stimuli are controlled in such a way as to encourage an equal number of target errors across IP positions. The result is taken to confirm the hypothesis that the IP serves as a planning domain.

6.4. Conclusions

In summary, higher error rates were observed when a pair of error-prone syllables straddled an AP boundary than when a pair of error-prone syllables occurred within a single AP. This pattern of results was even stronger when the distance between error-prone syllables was controlled. Given that the Across-AP tongue twisters were designed so that the error prone syllables shared AP position, the result that a greater number of target errors were elicited in these tongue twisters indicates a strong tendency towards structure preservation. Together, the findings support the idea that the Korean AP provides a prosodic (tonal) frame for phonological planning and that this frame is embedded in an IP, which is activated as a whole. The model proposed to account for the findings is as in Figure 6.1, panel (a). All syllables within the domain of the IP are retrieved and held in working memory while being slotted into their designated positions within the APs. Given that the tonal status of AP-initial syllables interacts with the identity of the onset segment, the further proposal (advanced in Chapter V) is that

syllables tagged for initial position may be held in reserve and slotted (from left-to-right)
only after other positions within the AP have been filled.

CHAPTER VII

CONCLUSIONS AND FUTURE DIRECTIONS

The current dissertation investigated the distribution of speech errors within various supralexical prosodic units in two different populations of speakers, adults and school-aged children, and in two languages, English and Korean. This chapter presents a summary of the main findings and the conclusions that were drawn regarding the psychological reality and the role of high- and mid-level supralexical prosodic units in speech planning and production. In addition, directions for future research are suggested based on limitations in the design and in the ecological validity of the experiments presented in the preceding chapters.

7.1. Distribution of speech errors in supralexical prosodic units

The results reported in Chapters III through VI all indicate that speech errors are systematically distributed within supralexical prosodic units (i.e., IPs and APs). Like the widely studied distributional patterns of errors in word-level linguistic units (i.e., the syllable and lexical word), the systematic distribution of errors within supralexical prosodic units provides us with evidence that these units are psychologically real and are functional units of speech planning.

Speech errors were found to be gradiently distributed within the largest supralexical prosodic unit, the Intonational Phrase (IP), in child and adult speech in English and in adult speech in Korean. The fewest speech errors occurred in IP-initial position with the number of errors gradually increasing across the unit. The gradient

increase in errors was referred to as the cumulative error pattern. We argued that this pattern implies the overall activation of an IP-sized unit during planning. Activation of this largest unit naturally decays over time, resulting in lower levels of activation in the subordinate units that are encoded towards the end (right-edge) of the IP.

The distributional error pattern within an IP is also consistent with the idea that a subsequent IP may be readied for encoding before the process is completed in a current IP (see also Krivokapić, 2007, 2012). In particular, we found that, when an IP is sentence internal (i.e., its right edge does not correspond to a sentence boundary), a large proportion of errors that occur in final position are anticipatory. This pattern was interpreted to suggest that elements in a subsequent IP are activated and can interfere with the planning and production processes in the current IP.

In contrast to the distribution of errors within IPs, there was a more categorical distribution of speech errors within the mid-level supralexical prosodic unit investigated, namely, the Accentual Phrase (AP) in Korean. In particular, a disproportionate number of errors occurred in AP-initial position: 59.39% of all errors occurred in the first syllable of APs, which had a mean length of 3.8 syllables (Chapter V). Moreover, errors were distributed fairly evenly across the remainder of the syllables in the AP: 20.34% occurred on the final syllable in the AP, leaving 20.27% to be distributed across the remaining 1 or 2 syllables. Thus, the pattern that emerges from these data corresponds best with the pattern illustrated in panel (a) of Figure 3.1, which was inspired by descriptions of the distribution of errors within the word in English. In addition, the results presented in Chapter VI indicated a pattern of structure preservation at the level of the AP, such that errors were much more frequent in error prone syllables when these shared the same

position in two different APs than when they occurred within a single AP, and so could not share an AP position.

Together, the results on the distribution of errors within APs strongly suggests that the AP functions as a structural unit during planning, providing a frame for the serial ordering of elements. We might also add that the strong evidence for a mid-level prosodic frame in Korean suggests to us that the relevant mid-level frame in English is more likely the metrical foot than the lexical word; that is, the identity of the English frame is likely more akin to that which is suggested in Shattuck-Hufnagel's (1983, 1992) slot-and-filler models than to that which is suggested in Dell's (Dell, 1986; Dell et al., 1997) spreading activation models. Caveats to our interpretation of the AP as a prosodic planning frame are presented below.

7.2. Planning domains versus planning frames

The assumed link between working memory and planning (Chapter I) is supported by the finding in Chapter III that working memory capacity predicts a higher proportion of anticipatory to perseveratory errors in children's speech. The relationship between children's verbal working memory capacity and IPs length must also therefore be interpreted to support the hypothesis that the IP is relevant to speech planning. The finding that working memory scores were inversely related to length was unexpected, but is nonetheless consistent with the idea developed from the gradient distribution of errors within an IP, namely, that the IP represents a planning domain. In particular, the result was interpreted to suggest that children with better verbal working memories chunk their speech more thoroughly than children with poorer verbal working memories, perhaps as a

strategy for minimizing confusions that arise from the frequent repetition of an element. In this section, we clarify the notion of a planning domain and contrast it with the notion of a planning frame.

A domain is to be understood differently than a frame. Whereas a frame has an internal structure into which elements are slotted, a domain is where a single operation is applied equally across constituent elements within its bounds. As suggested many times throughout this dissertation, the gradient increase in errors across the IP may be understood in terms of a single event applied to the unit as a whole: activation. The immediate consequence of activation is decay, and this has been our explanation for why during the left-to-right encoding process, errors increase across the domain. There are several phonetic patterns that also suggest that decay processes are relevant to speech production. Consider, for example, declination. Acoustic and articulatory studies have long noted that specific phonetic characteristics such as fundamental frequency, F1 and F2 height, lip and jaw opening, and velic raising are all higher or more robust in utterance-initial segments than in utterance-final segments, and that a decline in frequency or weakening in articulatory gestures is gradual across the utterance (Vatikiotis-Bateson & Fowler, 1988; Vayra & Fowler, 1992; Krakow, Bell-Berti, & Wang, 1994). Similar to what we have argued here, these effects have been interpreted to suggest that the level of activation associated with articulation slowly decreases over the span of an utterance. If we assume that the definition of an utterance in these acoustic and articulatory studies corresponds with our definition of an IP (a reasonable assumption given the length of most stimuli in these experiments), then these data argue for a decay

process that specifically occurs during the stage in planning when articulatory goals are being specified and motor routines selected.

Also consistent with the idea of the IP as planning domain is the finding that the elicited errors in this dissertation tend to take their source from adjacent items within the same IP rather than from items in the same position in another IP. Of course, this finding could also be an artifact of our nearest-neighbor criterion for categorizing different types of contextual errors. But the explanation from artifact does not explain the case when the error and source were also frequently found within IPs that were produced as isolated utterances (e.g., *Greek grakes are great* for *Greek grapes*). The absence of IP structure preservation is also evident in error and source relationships at the edges of two IPs. Consider, for example, the error [*Don't pamper damp scamp cramp*]_{IP} [*that camp under ramp lamps*]_{IP} for [...*tramp*]_{IP} [*that camp...*]. Here, the source for *cramp* must be the subsequent *camp* (there are no other voiceless velar stops in the vicinity), thus the final element of the IP is anticipating a second element in the next IP. Together, these findings suggest that errors and sources generally do not preserve IP internal structure even though other linguistic structures (e.g., syllables) are clearly preserved.

The absence of structure preservation at the level of the IP suggests that IP-internal structure is not relevant to the planning process. As first discussed in Chapter II, this may be due to the fact that these IPs are highly flexible in length and in their specific phonological make-up. Note, however, that gradient error patterns and violations to structure preservation should not be taken to imply that the IP has no internal structure. Intonation contours associated with IPs put the lie to this idea in that these follow from pragmatic and syntactic structure (Cooper & Paccia-Cooper, 1980; Selkirk, 1984;

Hirschberg, 2004; Chong, 2012). Instead, we would simply argue that meaning-derived structure is irrelevant to serial ordering processes at the level of phonological encoding and speech motor planning.

Although the intrinsic structure of an IP may not guide the ordering, encoding, or spell-out processes, it is clear that units within the domain of an IP do. It has long been assumed that words and syllables provide planning frames in English. The results from this dissertation suggest that the AP provides this kind of planning frame in Korean. The assumption is of a hierarchical relationship between mid-level prosodic units (e.g., prosodic words and APs) and the IP. This assumption is supported by the finding that the number of errors associated with specific positions in the mid-level units is influenced by the position of the unit within the IP. Specifically, whether the unit is metrical foot or the AP, initial position is more error prone, and the number of errors associated with this position varies as a function of where this unit occurs within the IP.

An alternative account for the different distributional error patterns in the AP and IP invokes influences from lexical representations. The reason that APs were investigated in the present experiment is because it was thought that they could be more clearly differentiated from the lexical word than, say, the trochaic foot of English can be. However, it is clear from the tongue twister data (Chapter V) that APs are perhaps as confounded with stand alone conceptual units (words) in Korean as trochaic feet are with words in English. Thus, it is possible that the findings of an AP-initial effect and AP-structure preservation are attributable to the representation of lexical words rather than to the representation of a mid-level prosodic unit. Although we tried to minimize the possibility of a word effect in the experiment reported in Chapter VI by training speakers

on multi-word APs and by using experimentally manipulated nonsense tongue twisters, the nonsense tongue twisters were ultimately composed of fixed syllable strings in which only the initial error-prone syllables were different. Thus, the structure of the stimuli could have allowed participants to process each syllable string as a word-like chunk, rather than as syllables embedded in a prosodic frame.

If structure preservation and the AP-initial effect reflect the word-like chunking of nonsense stimuli rather than the independent psychological reality of the AP, then the effects might also be explained without reference to a planning frame. In particular, the effects could be attributed to the high activation of initial elements in stored units. Let us assume that serial ordering begins with the activation of an IP-sized plan and that this entails that a number of words (described by APs) are then retrieved and temporarily stored in working memory. Let us further assume that retrieval of the item depends mainly on the initial element within each of these subordinate units. This initial element is thus a landmark element. All subsequent elements in the unit follow from this first one, and are tightly linked to the first and to each other. In this case, the initial element is the most error prone because it is the most informative, and so the most active element of every sequence. Especially high activation within the IP domain would make these elements prone to mis-selection during serial ordering.

7.3. Future directions

The distributional patterns of speech errors in Korean (Chapters V and VI) clearly indicate that the highest-level supralexical prosodic unit (i.e., the IP) functions differently from the well-defined mid-level supralexical prosodic unit (i.e., the AP). In contrast, the

distributional patterns of errors within the big and small IUs of English (Chapters III and IV) indicate that these units function similarly to one another. Based on the similarity of error distributions within the big and small IUs and our juncture-dependent criterion for defining these units, we argued that big and small IUs were likely both instances of an IP, albeit with different relationships to the pause-delimited utterance. Thus, the present experiments provided a poor test of the hypothesis that mid-level supralexicical prosodic units are relevant to speech planning in English. In defense of this short-coming, we noted that there is substantial disagreement about the definition of mid-level units in theories of prosodic phonology. Nonetheless, a future direction for research would be to thoroughly investigate the relevance of these units for planning. One way in which this might be accomplished would be to adopt a more systematized convention of prosodic transcription (e.g., Mainstream American English Tones and Break Indices; Beckman & Hirschberg, 1994) in order to identify the boundaries of these units according to their theoretical definition.

Another avenue for future research is suggested by limitations in the use of tongue twisters for eliciting speech errors. Tongue twisters employ frequent alternations between similar sounds in the same structural position across words to confuse the speaker. Their effectiveness in doing so is why this methodology is widespread in the field of speech production. That said, it is reasonable to wonder whether the distributional patterns reported in this dissertation result specifically from the tongue twister task. For example, it is conceivable that the cumulative error pattern is the direct result of the alternating patterns that characterize tongue twisters. In particular, it could be that speakers make more errors toward the end of an IP because confusion increases towards

the end of a sentence as words with similar sounds are repeatedly accessed and articulated. This possibility cannot be discounted, but the finding that the cumulative error pattern recurred within IPs in sentences composed of more than one IP (in Chapters III and V) at least suggests that the confusion builds across the span of an IP, not across the span of an entire tongue twister. Although this finding is consistent with the IP as a planning domain, the cumulative error pattern should nonetheless be replicated using more natural speech materials and, ideally, natural speech corpora.

Again with respect to the use of tongue twisters, our explanation for the negative correlation between the standardized Nonword Repetition (NR) scores and the IP length in children's speech was that children with larger working memory capacities use prosody to chunk the intended utterances into smaller domains in an attempt to minimize serial ordering errors. The assumption is that children with better working memories are better able to marshal linguistic structure to plan speech than children with poorer working memories. Future work will aim to test this idea by investigating the relationship between working memory and prosodic structuring in speech elicitation tasks that use more natural speech stimuli (i.e., non-repetitive) that nonetheless vary in length and syntactic complexity. The prediction is that children with better working memory capacities will reproduce such sentences with more prosodic boundaries than children with poorer working memories, if indeed such children make better use of prosody during planning than their peers.

Finally, it must be acknowledged that the experimentally elicited speech in the current dissertation is quite different from speech that is spontaneously generated in everyday conversations. Not only do the decontextualized stimulus sentences of this and

other experimental studies represent a special kind of speech, but the manner in which they were presented and spoken may encourage a special kind of speech planning. Here as in most other experimental studies of speech production, adult participants were encouraged to read sentences that were orthographically written and child participants were given extensive practice with the speech that they would produce. Participants were also explicitly told to prepare their speech before they spoke and all participants repeated the same sentence several times over. It is possible that the longer preparation times and extensive practice encourage planning of longer stretches of speech than that which normally occurs in everyday conversations. In future work, this possibility could again be tested by examining error patterns in spoken language corpora.

APPENDIX A

ENGLISH TONGUE TWISTERS AND
MATCHED NONTONGUE TWISTERS

Asterisks indicate the tongue twisters used in Experiment 2.

I. Short Sentences

a. Tongue Twisters

- Please pay promptly.
- Greek grapes are great.
- Fat frogs fly past fast.
- Fred threw three free throws.
- Beth believes thieves seize skis.
- Tighten these knapsack straps.
- Sly Sam slurps Sally's soup.
- One smart fellow, he felt smart.
- Friendly Frank flips fine flapjacks.
- Check the Unique New York shop.

b. Non-Tongue Twisters

- Please pay on time.
- Greek food is great.
- Most frogs can jump high.
- Matt had two free throws.
- People think thieves are crooks.
- The blue knapsack was small.
- The dog slurped the yogurt.
- He's a jolly good fellow.
- Glenwood makes apple flapjacks.
- Check Trader Joes for cookies.

II. Medium Sentences

a. Tongue Twisters

- Cedar shingles should be shaved and saved.
- *This shop stocks socks with stripes and spots.
- Which wristwatches are Swiss wristwatches?
- The myth of Miss Muffet is famous.
- *Three grey geese in the green grass grazing.
- She sifted thistles through her thistle-sifter.
- Nine nice night nurses are nursing nicely.
- Six shimmering sharks sharply striking shins.

- Don't pamper damp scamp tramps that camp under ramp lamps.
- Lesser leather never weathered wetter weather better.

b. Non-Tongue Twisters

- We must save our environment.
- The socks he wears have dots and flowers.
- I never wear a fancy wristwatch.
- All the students should read the Greek myths.
- Geese are completely vegetarian.
- John brought thistles to the botany class.
- Nurses are responsible for the sick.
- This magazine has some striking photos.
- Students should take care of spirit lamps in the lab.
- Weather's played an important role in human history.

III. Long Sentences

a. Tongue Twisters

- *The crow flew over the river with a lump of raw liver.
- I correctly recollect Rebecca MacGregor's reckoning.
- *A big black bug bit a big black bear, making the big black bear bleed blood.
- I slit the sheet, the sheet I slit, and on the slitted sheet I sit.
- *Give papa a cup of proper coffee in a copper coffee cup.
- A skunk sat on a stump and thunk the stump stunk, but the stump thunk the skunk stunk.
- He would chuck, he would, as much as he could, and chuck as much wood as a woodchuck should.
- I saw Susie sitting in a shoe shine shop: where she sits she shines and where she shines she sits.
- *If Peter Piper picked a peck of pickled peppers, where's the peck of pickled peppers Peter Piper pickled?
- While we were walking, we were watching window washers wash Washington's windows with warm washing water.

b. Non-Tongue Twisters

- Cod liver oil is a great source for vitamin A and D.
- He recollected the story of Farmer John and the llama.
- Susie happily watched polar bears at the zoo sliding on ice.
- Slitting the cloth into strips, she sat prettily in the corner.
- Copper mines are depleted, so copper wiring's become expensive.
- She whispered a tune while sitting on the toilet in a noisy bathroom.
- I had a party last night, so I spent the day washing dishes and chucking trash.
- The sun was high in the sky and shining beautifully when the little boy went out walking.
- The people in this town came together to pickle vegetables in preparation for winter.
- The basic and important practice of washing hands before each meal is no longer enforced in schools.

APPENDIX B

KOREAN TONGUE TWISTERS

- *cheolsu chaegsang-eun cheolchaegsang*
- *hanyang yangjangjeom yeop han-yeong yangjangjeom*
- *sinin syangsong gasu-ui sinchun syangsong syo*
- *hanguggwangwanggongsa gwagjingwang gwangwanggwajang*
- *chil-wolchil-il-eun pyeong-yang chingu chinjeong chilsun janchisnal*
- *gyeongchalcheong cheolchangsal-eun oecheolchangsal-inya*
ssangcheolchangsal-inya
- *sangpyo but-in keun kkanotong-eun kkan kkanotong-inga an kkan kkanotong-inga*
- *jeogi-issneun malttug-i mal mael malttug-inya mal mos mael malttug-inya*
- *jeogi ganeun jeo sangjangsaga sae sang sangjangsanya heon sang*
sangjangsanya
- *yeopjib patjug-eun bulg-eun patjug-igo dwisjib kongjug-eun geom-eun*
kongjug-ida.
- *geomchalcheong soecheolchangsal-eun saeoecheolchangsal-inya*
heonsoecheolchangsal-inya
- *jeogi-issneun jeo ttwimteul-i naega ttwil ttwimteul-inga naega an ttwil*
ttwimteul-inga
- *jung-angcheong changsal-eun ssangchangsal-igo sicheong-ui changsal-eun*
oechangsal-ida.
- *golyeogo gyobog-eun gogeubgyobog-igo golyeogo gyobog-eun gogeub-*
wondan-eul sa-yonghaessda
- *naega geulin gilin geulim-eun mog-i gin gilin geulim-inga mog-i angin gilin*
geulim-inga
- *ganjang gongjang gongjangjang-eun janggongjangjang-igo doenjang*
gongjang gongjangjang-eun ganggongjangjang-ida.
- *bagbeomboggun-eun bambeojkkochnol-ileul gago bangbeobbog-yang-eun*
najbeojkkochnol-ileul ganda
- *jag-eun tokki tokkitong yeop keun tokki tokkitong keun tokki tokkitong yeop*
jag-eun tokki tokkitong
- *jeogigyesin jeobun-i bag beobhagbagsa-isigo yeogi gyesin ibun-i baeg*
beobhagbagsa-isida
- *meongmeong-ine kkulkkul-ineun meongmeonghaedo kkulkkulhago kkulkkul-*
ine meongmeong-ineun kkulkkulhaedo meongmeonghanda.
- *nega geulin gilin geulim-eun mos geulin gilin geulim-igo naega geulin gilin*
geulim-eun jal geulin gilin geulim-ida.
- *naega geulin guleum geulim-eun saeteolguleum geulin guleum geulim-igo,*
nega geulin guleum geulim-eun gisteol guleum geulin guleum geulim-ida.
- *gyeongchalcheong geomchalcheong oenjjog yulichang jung-ang soechangsal-*

*eun nog-i seun soechangsal-igo gyeongchalcheong geomchalcheong oleunjjog
yulichang jung-angcheolchangsal-eun nog-i anseun cheolchangsal-ida.*

- *deul-ui kongkkagjineun kkan kongkkagji-inga an kkan kongkkagji-inga? kkan
kongkkagjimyeon eotteohgo an kkan kongkkagjimyeon eotteonya? kkan
kongkkagjina an kkan kongkkagjina kongkkagjineun da kongkkagji-inde*
- *ap jib patjug-eun bulg-eun pat puspatjug-igo dwisjib kongjug-eun
haeskongdankong kongjug ulijib kkaejug-eun geom-eunkkae kkaejug-inde
salamdeul-eun haeskong dankong kongjug kkaejug jugmeoggileul silh-
eohadeola.*
- *an chogchoghan chokochib nala-e saldeon an chogchoghan chokochib-i
chogchoghan chokochib nala-ui chogchoghan chokochib-eul bogo
chogchoghan chokochib-i doego sip-eoseo chogchoghan chokochib nala-e
gassneunde chogchoghan chokochib nala-ui munjigiga neon chogchoghan
chokochib-i anigo an chogchoghan chokochib-inikka an chogchoghan
chokochib nala-eseo sal-a lago haeseo an chogchoghan chokochib-eun
chogchoghan chokochib-i doeneun geos-eul pogihago an chogchoghan
chokochib nalalo dol-agassda.*

REFERENCES CITED

- Allum, P. H., & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: The role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 791-810.
- Adams, A.-M., & Gathercole, S. E. (1995). Phonological working memory and speech production in pre-school children. *Journal of Speech and Hearing Research*, *38*, 403-414.
- Adams, A.-M., & Gathercole, S. E. (2000). Limitations in working memory: Implications for language development. *International Journal of Language & Communication Disorders*, *35*(1), 95-116.
- Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of verbal learning and verbal behavior*, *14*(4), 382-391.
- Baddeley, A. (1998). Working memory. *Comptes Rendus de l'Académie des Sciences-Series III-Sciences de la Vie*, *321*(2), 167-173.
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, *4*, 417-423.
- Beckman, M. E., & Hirschberg, J. (1994). *The ToBI Annotation Conventions*. Online MS. Available at http://www.ling.ohiostate.edu/~tobi/ame_tobi/annotation_conventions.html.
- Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology yearbook*, *3*(1), 5-70.
- Behrens, H., & Gut, U. (2005). The relationship between prosodic and syntactic organization in early multiword speech. *Journal of Child Language*, *32*(1), 1-34.
- Berg, T. (1991). Phonological processing in a syllable-timed language with pre-final stress: Evidence from Spanish speech error data. *Language and Cognitive Processes*, *6*, 265-301.
- Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*, 1-47.
- Bock, J. K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945-84). San Diego, CA: Academic Press.

- Bock, J., & Warren, R. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*, 47-67.
- Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of memory and language*, *31*(4), 429-460.
- Burgess, N., & Hitch, G. J. (1996). A connectionist model of STM for serial order. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 51-72). Psychology Press.
- Butterworth, B. (1980). *Language production (Vol. 1)*. New York: Academic Press.
- Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2), 149-180.
- Chen, A., & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (Saarbrücken)* (pp. 315-320).
- Chen, J.-Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied Psycholinguistics*, *20*, 289-301.
- Chen, J.-Y. (2000). Syllable errors from naturalistic slips of the tongue in Mandarin Chinese. *Psychologia*, *43*, 15-26.
- Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, *46*, 751-781.
- Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a,i/ in English. *Journal of the Acoustical Society of America*, *117*(6), 3867-3878.
- Cho, T. (2006). Manifestation of prosodic structure in articulation: Evidence from lip kinematics in English. In L. M. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory phonology 8: Varieties of phonological competence*. Berlin & New York: Mouton de Gruyter.
- Cho, T., & Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics*, *29*, 155-190.
- Choe, W.K. (2011). The distribution of speech errors in prosodic phrases in Korean. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS-11 Hong Kong)* (pp. 480-483).
- Choe, W.K., & Redford M.A. (2012). The distribution of speech errors in multi-word prosodic units. *Laboratory Phonology*, *3*, 5-26.

- Chong, A. J. (2012). A preliminary model of Singaporean English intonational Phonology. *UCLA Working Papers in Phonetics*, 111, 41-62.
- Cooper, W., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.
- Croot, K., Au, C., & Harper, A. (2010). Prosodic structure and tongue twister errors. In C. Fougeron, B. Kühnert, M. D'Imperio & N. Vallée (Eds.), *Laboratory phonology 10* (pp. 433-460). Berlin & New York: Mouton de Gruyter.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of memory and language*, 25(4), 385-400.
- Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321.
- Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological review*, 104(1), 123-147.
- Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, 17(2), 149-195.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test, (PPVT-4)*. Minneapolis, MN: Pearson Assessments.
- Fay, D. (1982). Substitutions and splices: a study of sentence blends. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 163-195). Walter de Gruyter.
- Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological review*, 100(2), 233-253.
- Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101(6), 3728-3740.
- Fowler, C. (2007). Speech production. In M. G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 489-501). Oxford University Press.
- Frisch, S. (2000). Temporally organized lexical representations as phonological units. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the lexicon* (pp. 283-298). Cambridge University Press.
- Fromkin, V. (Ed.). (1973). *Speech errors as linguistic evidence (No. 77)*. Walter de Gruyter.

- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27-52.
- Gallo, C. G., Jaeger, T. F., & Smyth, R. (2008). Incremental syntactic planning across clauses. In B. C. Love, K. McRae & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (pp. 1294-1299). Austin, TX: Cognitive Science Society.
- García-Albea, J. E., del Viso, S., & Igoa, J. M. (1989). Movement errors and levels of processing in sentence production. *Journal of Psycholinguistic Research*, 18, 145–161.
- Garrett, M. F. (1976). The Analysis of Sentence Production. In G. H. Bower (Ed.), *The psychology of learning and motivation*, vol. 9 (pp. 133-177). New York: Academic.
- Garrett, M. F. (1980a). Levels of processing in sentence production. *Language production*, 1, 177-220.
- Garrett, M. F. (1980b). The limits of accommodation: Arguments for independent processing levels in sentence production. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 263-271). Academic Press.
- Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive psychology*, 15(4), 411-458.
- Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, 21(6), 649-683.
- Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, 103(3), 386-412.
- Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105, 611–633.
- Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*, 35(1), 1-31.
- Hayes, B. (1989). The prosodic hierarchy in meter. *Phonetics and phonology*, 1, 201-260.
- Hirschberg, J. (2004). Pragmatics and intonation. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 515-537). Oxford: Blackwell.

- Hokkanen, T. (2001). *Slips of the tongue: errors, repairs, and a model (Vol. 10)*. Finnish Literature Society.
- Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish & M. Zock (Eds.), *Current research in natural language generation* (pp. 287-319). London: Academic Press.
- Jaeger, J. J. (1992). Phonetic features in young children's slips of the tongue. *Language and Speech*, 35, 189-205.
- Jun, S.-A., & Fougeron, C. (1995). The accentual phrase and the prosodic structure of French. In *Proceedings of the 13th International Congress of Phonetic Sciences* (Vol. 2) (pp. 722-725).
- Jun, S.-A., & Fougeron, C. (2000) A phonological model of French intonation. In A. Botinis (Ed.), *Intonation: analysis, modeling and technology* (pp. 209-242). Dordrecht: Kluwer Academic Publishers.
- Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody*. Ph.D. dissertation, Ohio State University.
- Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, 15(2), 189-226.
- Jun, S.-A. (2000). *K-ToBI (Korean ToBI) Labeling Conventions*, <http://www.linguistics.ucla.edu/people/jun/ktobi/K-tobi.html>.
- Jun, S.-A. (2005a). Intonational phonology of Seoul Korean revisited. *UCLA Working Papers in Phonetics*, 104, 14-25.
- Jun, S.-A. (2005b). Korean intonational phonology and prosodic transcription. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 201-229). Oxford: OUP.
- Jun, S.-A. (2005c). Prosodic typology. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 430-458). Oxford: OUP.
- Keating, P. A. (2006). Phonetic encoding of prosodic structure. In J. Harrington & M. Tabain (Eds.), *Speech production: Models, phonetic processes, and techniques* (pp. 167-186). Psychology Press.
- Keating, P. A., Cho, T., Fougeron, C., & Hsu, C. (2003). Domain-initial strengthening in four languages. In J. Local, R. Ogden & R. Temple (Eds.), *Papers in laboratory phonology VI: Phonetic interpretations* (pp. 145-163). Cambridge, UK: Cambridge University Press.

- Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA working papers in phonetics*, 101, 112-156.
- Kim, S. 2004. *The Role of Prosodic Phrasing in Korean Word Segmentation*. Doctoral dissertation, UCLA.
- Klapp, S. T., Greim, D. M., & Marshburn, E. A. (1981). Articulation and the articulatory loop: Two names for the same mechanism or two distinct components of short-term memory? In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 459-472). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Krakow, R. A., Bell-Berti, F., & Wang, Q. E. (1995). Supralaryngeal declination: evidence from the velum. In R. Bell-Berti (Ed.), *Producing speech: contemporary issues for Katherine Safford Harris* (pp. 333-353). Springer.
- Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of phonetics*, 35(2), 162-179.
- Krivokapić, J. (2012). Prosodic planning in speech production. In S. Fuchs, M. Wehrich, D. Pape & P. Perrier (Eds.) *Speech planning and dynamics* (pp. 157-190). Frankfurt am Main, Germany: Peter Lang.
- Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.
- Lee, H. Y. (1990). *The structure of Korean prosody*. Doctoral dissertation, University of London.
- Levelt, W. J. (1989). *Speaking: From intention to articulation* (Vol. 1). MIT press.
- Levelt, W. J. (1999). Models of word production. *Trends in cognitive sciences*, 3(6), 223-232.
- Levelt, W. J.M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1-75.
- Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of verbal learning and verbal behavior*, 20(3), 298-305.
- Meyer, A. S., & Belke, E. (2007). Word form retrieval in language production. In M. G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 471-487). Oxford University Press.
- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

- O'Seaghdha, P. G., & Chen, J. Y. (2009). Toward a language-general account of word production: The proximate units principle. In *Proceedings of the 31st annual conference of the Cognitive Science Society* (pp. 68-73).
- O'Seaghdha, P. G., Chen, J. Y., & Chen, T. M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, *115*(2), 282-302.
- Park, M.-J. (2013). *The meaning of Korean prosodic boundary tones*. Leiden: Brill.
- Pérez, E., Santiago, J., Palma, A., & O'Seaghdha, P. G. (2007). Perceptual bias in speech error data collection: Insights from Spanish speech errors. *Journal of psycholinguistic research*, *36*(3), 207-235.
- Perrier, P., & Fuchs, S. (to appear). Motor equivalence in speech production. In M. A. Redford (Ed.), *The handbook of speech production*. Cambridge, MA: Wiley-Blackwell.
- Perrier, P. (2012). Gesture planning integrating knowledge of the motor plant's dynamics: A literature review from motor control and speech motor control. In S. Fuch, M. Weirich, D. Pape & P. Perrier (Eds.), *Speech Planning and Dynamics* (pp. 191-238). Frankfurt am Main, Germany: Peter Lang Publishers.
- Pfau, R. (2009). *Grammar as processor: a distributed morphology account of spontaneous speech errors* (Vol. 137). John Benjamins Publishing.
- Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 90–117). Cambridge, UK: Cambridge University Press.
- Prat-Sala, M., & Branigan, H. (2000). Discourse constraints on syntactic processing in language production: A crosslinguistic study in English and Spanish. *Journal of Memory and Language*, *42*, 168-182.
- Prieto, P., & Vanrell, M. M. (2007). Early intonational development in Catalan. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*(Saarbrücken) (pp. 309-314).
- Redford, M.A. (to appear). The acquisition of temporal patterns. In M. A. Redford (Ed.), *The handbook of speech production*. Cambridge, MA: Wiley-Blackwell.
- Roelofs, A. (1996). Morpheme frequency in speech production: Testing WEAVER. In G.E. Booij & J. van Marle (Eds.), *Yearbook of morphology 1996* (pp. 135-154). Dordrecht: Kluwer Academic Publishers.

- Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, 64(3), 249-284.
- Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes*, 14(2), 173-200.
- Schnur, T. T., Costa, A., & Caramazza, A. (2006). Planning at the phonological level during sentence production. *Journal of psycholinguistic research*, 35(2), 189-213.
- Schwartz, M. F., Saffran, E. M., Bloch, D. E., & Dell, G. S. (1994). Disordered speech production in aphasic and normal speakers. *Brain and language*, 47(1), 52-88.
- Selkirk, E. (1984). *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge: MIT Press.
- Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology yearbook*, 3, 371-405.
- Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 550-569). Cambridge, MA: Blackwell.
- Shallice, T., & Butterworth, B. (1977). Short-term memory impairment and spontaneous speech. *Neuropsychologia*, 15, 729-735.
- Shattuck, S. R. (1975). *Speech errors and sentence production*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.
- Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial order mechanism in sentence processing. In W. E. Cooper & E. C. Walker (Eds.), *Sentence processing: psycholinguistic studies presented to Merrill Garrett* (pp. 295-342). Hillsdale: Lawrence Erlbaum Associates.
- Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 109-136). New York: Springer.
- Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language*. Hillsdale, NJ: Erlbaum.
- Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, 42(1), 213-259.

- Shattuck-Hufnagel, S. (to appear). Prosodic frames in speech production. In M. A. Redford (Ed.), *The handbook of speech production*. Cambridge, MA: Wiley-Blackwell.
- Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of psycholinguistic research*, 25(2), 193-247.
- Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech, Language and Hearing Research*, 37(4), 831.
- Solomon, E. S., & Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology*, 49(1), 1-46.
- Stemberger, J. P. (1985). *The lexicon in a model of language production*. New York, NY: Garland.
- Stemberger, J. P. (1989). Speech errors in early child language production. *Journal of Memory and Language*, 28(2), 164-188.
- Stemberger, J. P. (1991). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language*, 30(2), 161-185.
- Stemberger, J. P. (2009). Preventing perseveration in language production. *Language and Cognitive Processes*, 24(10), 1431-1470.
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Approach to Child Language Acquisition*. Cambridge, MA: Harvard University Press.
- Tseng, C. H. (1997). Syllable as the segmentation unit in Chinese: New evidence. *Proceedings of the National Science Council, Part C: Humanities and Social Sciences*, 7(4), 612-631.
- Tseng, C. H., & Tsao, F. M. (1994). The gap inferiority effect in identifying Mandarin-Chinese phonemes. *Journal of National Chung Cheng University, Sec. II: Social Sciences*, 5(1), 181-204.
- Tseng, C. H., Huang, K. Y., & Jeng, J. Y. (1996). The role of the syllable in perceiving spoken Chinese. *Proceedings of the National Science Council*, 6(1), 71-86.
- Van der Merwe, A. (1997). A theoretical framework for the characterization of pathological speech sensorimotor control. *Clinical management of sensorimotor speech disorders*, 2, 3-18.

- Vatikiotis-Bateson, E., & Fowler, C. A. (1988). Kinematic analysis of articulatory declination. *Journal of the Acoustical Society of America*, 84, S128.
- Vayra, M., & Fowler, C. A. (1992). Declination of supralaryngeal gestures in spoken Italian. *Phonetica*, 49(1), 48-60.
- Vousden, J. I., & Maylor, E. A. (2006). Speech errors across the lifespan. *Language and cognitive processes*, 21, 48-77.
- Vousden, J. I., Brown, G. D., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*, 41(2), 101-175.
- Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and cognitive processes*, 25, 905-945.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing: CTOPP*. Pro-Ed.
- Warren, H. (1986). Slips of the tongue in very young children. *Journal of psycholinguistic research*, 15(4), 309-344.
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91, 1707-1717.
- Wijnen, F. (1992). Incidental word and sound errors in young speakers. *Journal of Memory and Language*, 31(6), 734-755.