CORPUS-BASED TRANSITIVITY BIASES IN PEOPLE WITH APHASIA

By

JENNIFER NICOLE DILALLO

_____

A Thesis Submitted to The Honors College
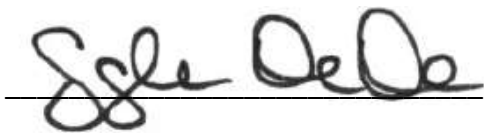
In Partial Fulfillment of the Bachelors degree
With Honors in

Speech, Language, and Hearing Sciences

THE UNIVERSITY OF ARIZONA

D E C E M B E R 2 0 1 5

Approved by:

_____

Dr. Gayle DeDe
Department of Speech, Language, & Hearing Sciences

Abstract

This study investigated whether people with aphasia (PWA) retain verb transitivity biases in expressive language. Transitivity, which represents a rudimentary division among simple sentence structures, is a fundamental language feature for our characterization of people with aphasia's (PWA) basic grammatical profile. However, no studies have yet detailed PWA's transitivity biases, given limited sample sizes of PWA (which especially restrict analyses of spontaneous speech). The current study analyzed 236 transcribed interviews of PWA from *AphasiaBank* and found that that PWA demonstrated identical transitive biases to controls. Further, PWA produced more intransitive verbs than transitive verbs overall. In ungrammatical productions, PWA's error rates were both higher in sentence structures that conflicted with verb bias and highest when an intransitive verb was attempted in a transitive structure. Thus, these findings indicate that PWA are sensitive to verb bias and verb complexity within expressive language. These effects are consistent with previous literature concerning PWA's sensitivity to verb bias in receptive tasks and to verb complexity in verb retrieval tasks.

Keywords: aphasia, transitivity, verb bias, expressive language, AphasiaBank

Corpus-Based Transitivity Biases in People with Aphasia

Our ability to better treat people with aphasia (PWA) largely depends on specifying knowledge about how aphasia affects language faculties. *Aphasia* encompasses a number of language impairments resulting from brain damage, usually due to stroke. In the United States, approximately 1 in 250 people suffer from aphasia (ASHA, n.d.), and the majority of PWA are adults above the age of 65 (Engelter et al., 2006). Studies of aphasia typically recruit PWA through rehabilitative groups within their community, and are often limited by population size and the recovery process (eg. participant exhaustion, different recovery patterns, mutual communicability, etc.). Thus, our current understanding of aphasia is largely based on small, variable samples of participant groups.

Further, aphasia is a highly variable condition. It may impair any combination of language domains, including verbal expression, auditory comprehension, reading comprehension, and writing. To better understand the nuances of linguistic deficits in aphasia, researchers specify a *feature* of language within a given *domain* to investigate whether and how aphasia changes normal language processing. A better understanding of how language breaks down in PWA may lead to better approaches to the evaluation and treatment of communication impairments in this population.

One fundamental language feature is *verb bias*, which refers to the frequency with which verbs occur in different sentence structures. Here, we focus on verb transitivity biases. *Transitivity* refers to the presence of a direct object in sentence

structures, such as in examples (1) and (2). Structures that do not include a direct object

are considered *intransitive* (examples (3) and (4)).

     1. I <u>dance</u> [*D.O.* the tango].

     2. I <u>eat</u> [*D.O.* tacos] [*Adv* aggressively].

     3. I <u>dance</u>.

     4. I <u>eat</u> [*Adv* aggressively].

The presence or absence of a direct object is important for describing a verb's

underlying *argument structure*. Transitive structures have at least two verb arguments

(e.g., agent and patient), which take the grammatical roles of subject and direct object

in a sentence, respectively. Intransitive structures have only one verb argument (e.g.,

agent), which takes the grammatical role of subject in the sentence. Given the

difference in the number of arguments, some studies consider the *verb* and/or

structural complexity of transitive versus intransitive structures. In this case, simple

transitive structures may be considered more complex than simple intransitive

structures since they have an additional argument (the direct object). Some verbs have

obligatory argument structures, meaning that they can only occur with a direct object or

without one. Examples (1) – (4) exemplify verbs with multiple argument structures. Such

verbs can occur in either transitive or intransitive structures.

The frequency with which a verb occurs in transitive versus intransitive

structures is referred to as that verb's *transitivity bias*. For instance, in unimpaired

speech, the verb "dance" occurs more frequently in intransitive sentences while the

verb "eat" occurs more frequently in transitive structures. There are many structures

other than transitive and intransitive frames (e.g. sentential complements, passives, and infinitives), and so there are multiple kinds of verb bias. For example, "know" is often followed by a sentential complement (e.g. "I know [*s.c.* that this is a sentential complement]"). Despite these alternative biases, simple transitive biases arguably capture of the most basic divisions among simple sentences in the English language. As we are interested in PWA's daily communication and fundamental representation of language, the current study focused on transitivity biases.

Transitivity biases affect both receptive and expressive language. Typically, receptive language encompasses reading and listening, while expressive language encompasses writing and speaking. In the context of verb biases, researchers first establish verb biases in expressive language by calculating the frequency with which a population produces a verb in different structures across large samples of written or spoken language. These biases then enable researchers to evaluate the effect of violations to those same biases in receptive language. That is to say, do these verb biases actually influence language processing? Taken together, this cross-domain approach holistically describes the influence of verb biases in language production and language comprehension for a given population.

Until now, no studies have described whether PWA show the same verb biases as controls in expressive language. However, previous studies have compared PWA's and controls' sensitivity to normative verb biases, and we can use that literature to develop predictions about the nature of PWA's verb biases in expressive language.

Literature Review

Transitive Biases in Expressive Language: Unimpaired Populations

First, let's look at existing transitive biases in control populations. Studies have determined verb biases through two main methods: sentence elicitation and corpus analysis. In elicitation studies, participants are provided with verbs and asked to generate one verb per sentence. In corpus studies, verbs are coded based on their usages in large language samples (e.g., taken from newspapers or structured interviews). The advantage of corpus databases is that they represent more naturalistic language productions, including information about the relative frequency of each verb regardless of bias. However, corpora samples may vary by sample context (eg. an academic journal vs. a radio station), population size, and coding decisions (eg. what kind of sentence structures to code and to what detail).

For this reason, cross-corpora studies like Gahl, Jurafsky, and Roland (2004) are especially helpful in understanding replicable trends across corpora. Their study focused on verb subcategorization frequencies in two separate corpora (the COMLEX database and British National Corpus). Gahl and colleagues developed a coding schema to label a number of sentential frames, including simple transitive and simple intransitive sentences. Then, they compared their verb biases to those of five elicitation-based studies. The team evaluated how various coding decisions affected verb biases, like the effect of using relative proportions of use or an absolute criterion (e.g. two thirds of all uses) to establish verb bias. For transitivity, the research team compared direct object (DO) versus sentential complement (SC) biased verbs, and found that 167 of 281 verbs

changed bias based on the analysis. Their study suggested that absolute criterions may overlook smaller, though significant, transitive biases. Gahl et al. (2004) then made publicly available their verb biases for the 281 verbs. Among their data, we can see that "dance" is used in intransitive structures 89.3% of the time, and in transitive structures 5.4% of the time; alternatively, "eat" is used in intransitive structures 18.9% of the time, and in transitive structures 67.8% of the time. Thus, with these informed analyses and verb biases, researchers can turn to receptive language to make predictions about how verb biases influence language processing.

Transitive Biases in Receptive Language: Unimpaired Populations and PWA

Within receptive literature, we can begin to examine PWA's sensitivity to verb bias, keeping in mind that current verb biases are normed on unimpaired populations. Logically, a fundamental question is if and how verb biases influence typical language comprehension. Then, we can ask the same question about aphasic language comprehension to generalize cross-group differences.

An early study by Garnsey, Pearlmutter, Myers, and Lotocky (1997) evaluated the relative influence of verb bias vs. plausibility in reading comprehension, while also comparing two methodologies: 1) eye-tracking and 2) self-paced reading. For both methodologies, the same sentences were presented to a group of college students (n=78 and n=81 for each method, respectively). The sentences contained three sets of verbs with different biases: one biased to direct objects (DO-bias), another to sentential complements (SC-bias), and a third with an equal preference for both structures (EQ-bias). These verbs appeared in structures with either a DO or SC, which represented

either a match or a mismatch with the verb's bias. The noun following the verb varied in

plausibility – either biasing the reader toward a DO- or SC-interpretation. Further, the

study manipulated the presence of the complementizer *that* in SC-structures, which

permitted investigators to analyze effects of ambiguity. Across both experiments,

Garnsey et al. (1997) found that verb bias mainly affected reading times in ambiguous

contexts where plausibility supported the alternative structural interpretation. That is,

participants read the disambiguating region of a sentence slower when verb bias

conflicted with the structure, and this effect was amplified when the plausibility of the

NP supported the first (incorrect) interpretation. This phenomenon suggests an

interaction between plausibility and verb bias in structural parsing. Further, Garnsey et

al. (1997) considered that the absence of plausibility effects on the NP following SC-bias

verbs may be an indicator for the early influence of verb bias. The research team

interpreted the variable reading times of the complementizer *that* (which were faster

for SC-bias verbs and slower for DO-bias verbs) as evidence for access to verb bias

before encountering the next word. Taken together, Garnsey et al. (1997) provided

support for the early influence of verb bias in structural parsing as separate from other

lexical factors like plausibility and *that*-preference. These effects were significant across

both experimental methodologies.

Staub (2007) also investigated how transitivity bias affects sentence

comprehension in unimpaired college age adults. To isolate the role of intransitivity bias

in structural parsing, Staub (2007) conducted a three-part eye-tracking study of

transitive and intransitive biases in sentences where a NP always followed the verb. In

his third experiment, 37 healthy adults read sentences with three verb types: NP-preference, PP-preference, and unaccusative verbs. Note that, despite verb preference, all verbs in the study permitted intransitive structures – and only the unaccusatives obligatorily. All verb types occurred in intransitive structures in either the relative clause or the main clause position (referred to as the gap v. no gap condition, respectively). In the gap condition, participants read both NP-preference and PP-preference with marked difficulty (that is, with a delay). Staub interpreted this finding to indicate that participants posited a NP after the verb regardless of its optional bias. However, participants did not replicate this trend with unaccusative verbs, which are strictly intransitive. This difference underscores the importance of the relative strength of verb bias in influencing structural parsing. Importantly, Staub's data revealed a significant effect of intransitivity in early parsing interpretations.

Studies of aphasic receptive language generally show that PWA are also sensitive to transitivity biases during language comprehension. Gahl (2002) performed a comprehension study of simple transitive, intransitive, and passive sentences with five controls and 17 PWA (with various aphasia types). The goal was to corroborate whether PWA were sensitive to lexical biases in sentence comprehension, known as the *Lexical Bias Hypothesis*. Participants listened to sentences with eight transitively-biased and eight intransitively-biased verbs in structures that either supported or violated the verb's biases (i.e. the *matched* and *mismatched* conditions, respectively). Further, the study alternated subject/object roles to create plausible and implausible conditions. To test their comprehension, participants made plausibility judgments at the conclusion of

each sentence. The control participants' error rates on this plausibility judgment task were higher in the mismatched condition when compared to the matched condition. Gahl (2002) explained this pattern in that the conflict of syntactic and semantic information made sentences (especially passives) more vulnerable to misinterpretation. Overall, PWA as a group replicated this trend, yet committed a proportionally greater amount of errors than controls on the mismatched transitive bias condition (i.e. a transitively biased verb in an intransitive sentence).

DeDe's (2013) study also considered the Lexical Bias Hypothesis in PWA by comparing the influence of verb bias versus structural ambiguity (via the presence/absence of the complementizer *that*) in online reading processing. Ten PWA and ten controls participated in a self-paced reading paradigm with sentences featuring sentential complements. Main verbs were biased either in favor of taking a direct object or sentential complement. Each sentential frame occurred with and without the complementizer *that.* Successful understanding of the sentence was measured through accuracy on a subsequent comprehension question. Despite slower reading times on the sentences and higher error rates on the questions, PWA's group performance did not significantly differ from controls in relation to sensitivity to verb bias. Interestingly, among individual data, a majority of PWA (seven of the nine participants) demonstrated a greater sensitivity to verb bias in comparison to sensitivity to structural ambiguity than controls. These findings suggest that PWA access verb bias to navigate sentential comprehension, and may indeed rely more on this lexical feature than control populations.

Across the receptive literature of both controls and PWA, the general trend is that both populations demonstrate effects of verb bias in language comprehension. In fact, it may be that PWA exhibit greater sensitivity to verb biases in receptive studies. However, until we establish verb biases specific to PWA and evaluate how those biases differ from those of controls, this discussion is limited to the influence of verb biases normed on populations without aphasia.

Transitive Biases in Expressive Language: People with Aphasia

To develop knowledge of verb biases in PWA, researchers need large samples of PWA's written or spoken language that include many uses of different verbs. Until recently, there was no corpus available to analyze the productive biases of PWA. Instead, researchers recorded spontaneous speech samples from small groups of PWA. Previous studies of these groups have considered effects of argument structure and verb phrase complexity on PWA's speech, though no study has directly examined verb bias in PWA's expressive language.

Saffran, Berndt, and Schwartz (1989) elicited speech from ten PWA to characterize their structural complexity while producing the Cinderella story (a common narrative in language research). Speech samples were transcribed and compared to five control participants. The coding schema included considerations for syntactic complexity, both in terms of accuracy (how well-formed sentences were) and elaboration of NP and VP constituents within a sentence. These measures were calculated separately, so it is unclear whether PWA made more errors on transitive or intransitive sentences. However, the authors reported that, of the 52 ill-formed

sentences in agrammatic samples, only two violations were due to a missing argument and four were due to both a missing argument and morphological issue. The structural errors committed by non-fluent participants were not described, but the authors mentioned that this group made less errors overall. Taken together, the results suggest that PWA do not typically omit arguments in sentences. The authors did not consider verb bias, however, and so we do not know how a verb's "preferred" argument structure might interact with argument omissions in PWA.

Webster, Franklin, and Howard (2007) found somewhat conflicting results in their analysis of spontaneous speech samples with PWA. This study included 22 PWA and 20 controls. Speech samples were coded using Whitworth's (1995) schema of thematic structure. This analysis permitted the research team to detail the omission of obligatory arguments in two and three argument structures. As a group, PWA produced more sentences with fewer or undetermined thematic structures than controls, and no embedding whatsoever. Interestingly, there was no significant difference between control populations and PWA in terms of verb phrase complexity. PWA were more likely to omit obligatory arguments than the control group, but it was not clear whether they made more errors on certain sentence types. The study did attempt to correlate overall sentence production with phrasal errors, and did not find significant trends between mean phrasal complexity and mean percentage of phrasal errors. Webster et al pointed to the high level of individual variability for the lack of group trends in error production. Overall, this study demonstrated both that PWA produce reduced sentence complexity

and that they are more likely to omit arguments. Again, this study did not examine

sentence structure as a function of verb bias.

Based on these studies, it seems that PWA produce less complex sentences and

may omit arguments – but we do not know in which structural contexts. We need to

specify these findings, especially in terms of verb bias, in order to better understand

PWA's basic grammatical profile.

The Current Study

It is unclear whether PWA exhibit the same verb biases in their speech

production as unimpaired populations. If verb biases are present in PWA's expressive

language, do violations to verb bias have any association to ungrammaticality? If verb

biases are not present in PWA's expressive language, are participants more sensitive to

intransitivity or transitivity overall? Such answers will nuance our representation of

PWA's language processing and production in order to improve current language

models and therapy practices.

A recent contribution to the field of aphasiology is the development of

AphasiaBank, a corpus database featuring speech samples of PWA from around the

United States (MacWhinney, Fromm, Forbes, & Holland, 2011). Participants with

different aphasia types (e.g., Broca's, Conduction, Anomic aphasia) generate

spontaneous speech samples in response to a range of prompts. Twenty-two

institutions contributed to the database following a shared protocol with a total of 324

unique participants (as of December 8, 2015). This contribution is monumental,

considering that it counters typical limitations of aphasic studies by providing data from

a large number of participants that are tested under the same methodology. Further, while it has been notoriously difficult to examine PWA's productive language through small languages samples, *AphasiaBank* provides a large enough database for researchers to begin describing PWA's productive language – including features as specific as different types of verb bias.

Thus, this study proposed to examine the presence and nature of transitive biases in PWA across *AphasiaBank*. In order to analyze both verb bias frequency and accuracy, we adapted the coding schema from Gahl et al. (2004) and proposed error codes to capture the nuances of aphasic language. We anticipated that, based on the strength of the Lexical Bias Hypothesis in PWA's receptive language, participants would holistically follow normative verb biases in their expressive language. Moreover, DeDe's (2013) findings gave us reason to postulate that PWA may have exaggerated representations of verb bias compared to controls. For example, intransitively-biased verbs may have relatively *greater* intransitive biases, and the converse for transitively-biased verbs. We also postulated that PWA's productions of mismatched structures (e.g. a transitively-biased verb in an intransitive structure) would be associated with higher error rates than matched structures.

However, if PWA do not retain sensitivity to verb bias, they may exhibit an overall preference for intransitive or transitive structures. A study by Kim and Thompson (2000) examined PWA's verb naming preferences as a function of argument complexity. To elicit target verbs, the research team asked seven participants with aphasia to name the actions of a given set of pictures. In this study, PWA produced more errors when

naming verbs with more arguments in comparison to verbs with one argument (i.e. the subject). If this preference for naming verbs with less complex argument structures influences PWA's speech, we would predict more intransitive productions than transitive productions (regardless of verb bias). In other words, PWA may favor constructions that are less *structurally complex*.

Alternatively, we might predict that PWA would favor transitive productions if they are most sensitive to *structural frequency*. Transitive productions represent typical canonical order in English (i.e. Subject-Verb-Object) and are produced more frequently by unimpaired populations (Gahl et al., 2004). If PWA have a preference for the structure that they most commonly interact with, they may produce more transitive structures. To clarify, these predictions are primarily relevant in the case that verb biases are not present in PWA's speech, but such considerations can inform whether PWA are more sensitive to structural complexity or frequency regardless.

Overall, the focus of this study was to characterize PWA's transitivity biases in language samples from *AphasiaBank*. We wanted to know if these biases look similar to control populations, and in what ways they may differ. We examined PWA's intransitive and transitive productions of a set of intransitively-biased and transitively-biased verbs (identified by Gahl et al, 2004's norms). We were also interested in whether PWA's error rates reflect conflicts between verb bias and structure, and the possibility that PWA exhibit an overall structural preference. Finally, we considered whether individuals with one aphasia type, Broca's aphasia, show a different pattern of verb biases than the group of controls. This was an important consideration because some authors (e.g., Kim

& Thompson, 2000) have argued that individuals with Broca's aphasia have more impaired verb processing than individuals with other aphasia types.

Methods

Participants

236 morphologically and grammatically transcribed interviews of participants with aphasia were analyzed from the *AphasiaBank* database using CHAT and CLAN transcription programs (MacWhinney, 2000). All participants were native speakers of English between the ages 25 and 90 years old. According to the Boston classification system, approximately 31% of participants were identified to have Anomic aphasia, 31% Broca's, 11% Conductive, 6% Global, 6% Wernicke's, and 3% transcortical motor (see Figure 1). The remaining 12% of participants' aphasia types were unidentified, non-classifiable, or other.

[Figure 1: Aphasia Types in *AphasiaBank*]

Materials

Participants in *AphasiaBank* performed the same tasks in a video-recorded session: 1) sharing their aphasia story and coping experiences, 2) talking about an important story event in their life, 3) narrating pictures, 4) narrating the Cinderella story, and 5) describing how to make a peanut butter sandwich. Instructions and examples of protocol are available on the *AphasiaBank* website. Interviews of PWA were downloaded October 31, 2014.

This study also includes 41 control participants from *AphasiaBank* who performed the same protocol as PWA. One modification is that these participants

shared any illness story for Task 1. These individuals ranged between 23-89 years, and were also native speakers of English. Interviews of controls were downloaded April 4, 2015.

Verbs of interest. 55 verbs were chosen for coding. The verbs were selected in part through an informal review of the transcripts in *AphasiaBank*. Our goal was to identify verbs that were likely to be elicited by the corpus' language tasks so as to establish relatively stable verb biases. We also matched verbs from our informal review with the list of verbs in Gahl et al. 2004's in order to make comparisons with the verbs' typical biases outside this study. The 55 selected verbs showed biases toward transitive, intransitive, passive, and sentential complement structures. This paper focuses on a subset of those verbs: 11 transitively-biased verbs and 11 intransitively-biased verbs. The 22 verbs were matched for frequency and number of syllables (Table 1).

[Table 1: Verb Tokens]

Coding. The coding scheme was based on the one described by Gahl et al. (2004) (see Appendix A for adjustments to this coding scheme). Two coders with training in linguistics applied the coding scheme to the verb tokens. Inter- and intra-rater reliability was measured through Cohen's Kappa (Қ), a statistic that measures percent agreement due to chance. Acceptable reliability was determined based on Gahl et al. (2004)'s criteria of a Kappa value greater than 0.80.

Reliability. Prior to data collection, the coders coded five transcripts, and had an inter-rater reliability of Қ = 0.77. The research team met to discuss coding conflicts, focusing on adaptations for aphasic speech, and reached coding decisions through

consensus. The second round of pre-coding (again with five transcripts) rendered an

inter-rater reliability of К = 0.83.

Throughout the coding process, the coders marked instances of uncertainty to

discuss in lab meetings. Coding decisions for these utterances were decided on a

consensus-basis with the principal investigator. The average inter-rater reliability

following coding was К = 0.86 (and this value was the same for both coders). For this

calculation, each coder re-coded approximately 8% of the other coder's files (based on

total number of target verb uses). For intra-rater reliability, each coder re-coded 5% of

their own files, and both coders were above the acceptability criteria (К = 0.92 and К =

0.88).

Results

Verb Biases

For each verb, the proportion of transitive or intransitive uses was calculated by

dividing the number of transitive or intransitive uses by the total number of occurrences

for the verb. This proportion was the dependent variable in 2 (group: PWA vs. controls)

by 2 (verb bias: transitive vs. intransitive) ANOVAs, separately for the proportion of

intransitive uses and the proportion of transitive uses.

[Figure 2: Transitive Verb Uses]

Proportion of Transitive Uses. Figure 2 presents how PWA and controls used

transitive verbs in the Aphasia Bank sample. For transitive structures, transitive verbs

accounted for an average of 48.39% of uses in PWA (eg. *I eat tacos aggressively*).

Intransitive verbs were used in transitive structures 2.06% of the time in PWA (eg. *I*

dance the tango). For controls' transitive productions, transitive verbs accounted for

51.42% of these uses while intransitive verbs contributed 5.73% of productions. The

main effect of verb bias was significant, F1 (1,40) = 68.39, p< .001. The main effect of

verb bias was due to an increased proportion of transitive uses for transitively-biased

verbs, for both groups. Neither the main effect of group nor the interaction of group

and verb bias approached significance (F's < 1).

[Figure 3: Intransitive Verb Uses]

Proportion of Intransitive Uses. Figure 3 presents how PWA and controls used

the intransitive verbs. Both groups also produced higher proportions of intransitive

structures with intransitive verbs. In PWA, intransitive verbs accounted for 62.75% of

intransitive productions (eg. *I dance*), compared to 17.52% with transitive verbs (eg. *I

eat aggressively*). Controls similarly produced intransitive sentences with a higher

proportion of intransitive verbs (60.75%) compared to transitive verbs (11.68%). The

main effect of verb bias was significant, F(1,40)=29.64, p<.001. The main effect of group

and the interaction of group by verb bias were not significant (F's <1).

[Figure 4: Overall Verb Preference]

Intransitive or Transitive Verb Preference

Even though PWA and controls showed similar verb biases, we were interested

in determining which set of verbs PWA produced more frequently (i.e. intransitively-

biased or transitively-biased verbs). It is possible that PWA would show a preference for

transitive or intransitive verbs when speaking, even while maintaining the same verb

biases as controls. Thus, the question remains whether PWA are more sensitive to verb

complexity (which would favor production of intransitive verbs) or the overall frequency

of SVO structures (which would favor the production of transitive verbs). To address

this question, we analyzed the total number of times that PWA and controls used the

transitive and intransitive verbs, regardless of sentential structure.

As Figure 4 shows, both PWA and control populations produced more

intransitive than transitive verbs (regardless of sentential structure). However, the

preference for producing intransitive verbs was larger in PWA than controls. (Chi-square

(1)=27.8, p<.001). PWA produced 2174 intransitive verbs (62.5% of all productions), as

compared to 1307 transitive verbs (37.5% of all productions). Controls produced 752

intransitively-biased verbs (54.2% of all productions), as compared to 635 transitively-

biased verbs (45.8% of all productions).

Aphasia Type

Based on previous literature on agrammatism, we were interested in considering

the transitive bias of participants with Broca's aphasia as separate from the larger PWA

group. Some authors have argued that individuals with agrammatic Broca's aphasia

have more impaired verb retrieval than other aphasia types. Although this argument has

not been made with respect to verb bias, it has been made with respect to sensitivity to

verb complexity (Kim & Thompson, 2000). To investigate this issue, we examined

whether the subset of individuals with Broca's aphasia showed a different pattern of

transitivity biases than the group as a whole. There was a significant effect of verb bias

for both the proportion of transitive ($F1$ (1,20) = 19.38; p<0.001) and intransitive ($F1$

(1,20) = 12.93; p<0.002) uses. In transitive uses, transitive verbs accounted for 46.83% of

transitive productions, as compared to 3.74% of transitive structures produced with

intransitive verbs. In intransitive uses, intransitive verbs accounted for 70.81% of

productions, and transitive verbs accounted for 26.54%. These results suggest that verb

biases for individuals with Broca's aphasia were similar to both the controls and the

larger group of PWA.

[Figure 5: Error Rates]

Error Rates

These analyses evaluate PWA's ungrammatical productions in terms of whether

there were more errors when verb bias conflicted with sentence structure. That is, are

PWA more likely to produce errors when using transitively-biased verbs in intransitive

vs. transitive sentences? Error rates were analyzed in 2 (Verb bias) x 2 (Sentence

Structure) ANOVAs. The proportion of utterances with errors is presented in Figure 5.

The main effect of sentence structure was not significant, $F_{(1,40)} = 3.5$, $p=.07$. There

was a main effect of verb bias, $F_{(1,40)} = 4.7$, $p=.04$. PWA made more errors in the

mismatched than matched condition overall. Interestingly, the interaction of verb bias

and structure $F_{(1,40)} =10.4$, $p<0.01$ was also significant. PWA's errors on intransitive

structures with a transitively-biased verb (i.e. the mismatched condition) accounted for

50.48% of all errors, while intransitive structures with an intransitively-biased verb (i.e.

the matched condition) accounted for 9.87% of all errors. Error rates in transitive

structures followed the same pattern. In the mismatched condition for a transitive

structure (i.e. an intransitively-biased verb), PWA committed 18.33% of all errors, and

7.56% in the matched condition. Thus, PWA made more errors when using

intransitively-biased verbs in transitive structures than when using transitively-biased verbs in intransitive structures.

Discussion

PWA preserve sensitivity to verb bias in their spontaneous speech. We do not see group differences between PWA and controls for the transitive or intransitive verbs investigated in this study. Both controls and PWA produce intransitive sentences proportionately more with intransitively-biased verbs, and transitive sentences proportionately with more transitively-biased verbs. The findings are consistent Gahl et al. (2004)'s verb biases for both PWA and control populations. The results are also consistent with previous studies (DeDe, 2013; Gahl, 2003), which demonstrated that PWA are sensitive to verb biases in sentence comprehension

PWA also exhibited a preference for intransitive verbs. Overall, they produced more intransitive verbs than transitive verbs – regardless of sentential structure – and this difference (though in the same direction) was greater than that of controls. This preference supports findings from Kim and Thompson (2000), who reported that PWA named obligatory one-place and optional two-place verbs (i.e. obligatory intransitive verbs and optional (in)transitive verbs) more accurately than obligatory two-place verbs (i.e. obligatory transitive verbs). The idea is that intransitive verb argument structures are less complex than transitive ones. This lower complexity may denote lower cognitive costs in sentence production, as participants do not have to retrieve an additional argument (i.e. the direct object). However, the findings from the present study do not

directly inform claims about cognitive cost – only the frequency of productions of intransitively-biased vs. transitively-biased verbs.

It is likely that this intransitive verb preference - together with transitivity biases - affect PWA's error rates. PWA made more errors on sentences with mismatched biases than those with matched biases. This finding indicates that violations of verb bias entail a higher likelihood of error. Of the two possible mismatched conditions (i.e., an intransitively-biased verb in a transitive structure, or a transitively-biased verb in an intransitive structure), PWA's attempts to produce intransitively-biased verbs in the opposite structure resulted in the highest error rates. We can specify, then, that violations of the preferred verb bias (i.e. an intransitive verb) are associated with higher error rates than violations of the less preferred verb bias – and this preference relates to verb complexity. Thus, analyses of PWA's ungrammatical productions demonstrate how sensitivity to verb bias and verb complexity can jointly influence the grammaticality of PWA's speech production.

To summarize: if PWA are more sensitive to verb complexity in language processing and, relatedly, prefer intransitively-biased verbs due to their higher probabilities of reduced complexity, it follows that utterances that violate this intransitive preference would signify higher cognitive costs for PWA. However, as we will see, this study cannot make direct claims about PWA's cognitive representation of transitivity.

Limitations

The present study compared verb biases in PWA and unimpaired controls. However, the process of coding PWA's speech samples introduced a range of possible confounds. First, our data inherited any confounds present in *AphasiaBank*. The participant population of PWA necessarily had to perform some, if not all, of the language tasks for *AphasiaBank*. Accordingly, there may be an underrepresentation of PWA with more severe language deficits or certain aphasia types that limit productive language abilities. Further, *AphasiaBank* is continually undergoing revision, and errors in the original transcript would influence our subsequent codes. We did not analyze clearly mislabeled items (e.g. gerunds as verbs and vice versa), though our analyses can be improved together with *AphasiaBank's* expansion and revisions.

Another concern is that the present study does not consider non-linguistic communication such as gestures, facial expressions, and pointing. It was beyond the scope of this study to consider how non-linguistic communication may fulfill grammatical roles, or to determine coding guidelines for them. For controls, this is unlikely to significantly affect data analyses. However, people with aphasia may represent simple nouns and verbs (which are crucial for determining basic sentence structure) through non-verbal, expressive communication.

Relatedly, we have to acknowledge the limits of our adapted coding scheme. One indicator of the scheme's scope is reliability. Our study saw high inter- and intra-rater reliability for this kind of research, especially considering that coders were grammatically labeling non-grammatical speech. Part of this reliability may be attributed

to our decision to avoid labeling uninterpretable utterances or abandoned utterances that seemed ambiguous. We also handled non-grammatical structures by labeling productions on a surface level. Such coding decisions are necessary, given that deeper level interpretations often lead to multiple possible grammatical possibilities. In these cases, there would be few objective guidelines to select one grammatical interpretation over another, all of which necessarily impose grammaticality on participants with disordered language processes.

However, one may question whether these decisions affect our study's validity: does the current coding scheme accurately represent PWA's transitivity usage? To discuss this issue, consider a few examples of ungrammatical productions:

5. *She <u>danced</u> the prince.

6. *I <u>think</u> progress, but I don't know.

7. *…because she <u>see</u> of something….

The current study labeled both structures (5) and (6) as transitive, but one can reasonably infer that the closest grammatical structures would involve different sentence frames: (5) an intransitive structure with the addition of "with," and (6) a sentential complement with the addition of "that there is". However, based on actual productions, we cannot assume that the participants did not intend to use either *dance* or "think" transitively in these cases.

Sentence (7) represents a grammatical violation away from transitivity. A more grammatical production would not include "of," resulting in a transitive structure (eg. "She sees something"). These structures exhibit that our coding decisions could not

account for structural intent, only production – and possible confounds occurred both in the direction of transitivity and intransitivity. Structure (7) also presents us with the consideration that multiple kinds of errors can exist within one utterance. Our study described errors based on position (before, on, or following the verb), and type (mainly inflectional or semantic). Even these basic labels proved difficult to tease apart, and we often collapsed across error type (i.e. *'error after verb'* could account for more than one error).

Altogether, though it would be highly relevant to describe ungrammatical productions in greater detail, our current scheme does not enable us to make finer distinctions of ungrammaticality with confidence. Since we do not have clear criteria for selecting one interpretation of an ungrammatical production above another, our study only captures the basic sentence structure and gross errors of surface-level productions. This limits our ability to directly examine PWA's cognitive representation of transitivity. At the same time, the current findings are consistent with existing literature on transitivity and verb bias for both unimpaired populations and PWA across expressive and receptive domains. This cross-study support lends credibility to the observed effects of verb bias and verb complexity in PWA's productive language. Future studies may continue improving both verb bias codes and error codes for aphasic speech to specify these effects in different and deeper contexts, especially as *AphasiaBank* continues to expand and undergo revisions.

Conclusion

This study characterized the presence and nature of verb biases in PWA's spontaneous speech, as recorded in structured interviews from *AphasiaBank*. Both controls and PWA exhibited verb biases in the same direction as normative verb biases from Gahl et al. (2004). There were no significant differences between PWA and controls' proportion of transitive or intransitive uses, indicating that PWA retained both simple transitive and intransitive verb biases. Overall, PWA (like controls) produced more utterances with intransitive verbs than with transitive verbs. This intransitive preference was greater for PWA. It may be that this verb preference belies PWA's sensitivity to verb complexity (Kim & Thompson, 2000), but this possibility will have to be further explored in future studies analyzing a range of verb complexities. PWA's error rates were correlated to verb bias; mismatched productions (between verb bias and sentence structure) had significantly higher error rates than matched productions. Of the mismatched conditions, PWA produced more errors when using intransitive verbs in transitive structures. This finding further supports the possibility of an intransitive preference. Thus, the data suggest that PWA retain typical simple transitive and intransitive verb biases, and demonstrate sensitivity to verb complexity in their expressive language. These initial characterizations of PWA's verb biases are consistent with existing literature in receptive domains. In future corpus-based studies, we want to consider differently biased verbs and continue refining our approaches to labeling PWA's ungrammatical speech.

References

American Speech-Language-Hearing Association. (n.d). *Aphasia: Incidence and Prevalence*.

DeDe, G. (2013). Effects of verb bias and syntactic ambiguity on reading in people with aphasia. *Aphasiology, 27*, 326-343.

Engelter, S. T., Gostynski, M., Papa, S., Frei, M., Born, C., Ajdacic-Gross, V., Gutzwiller, F., Lyrer, P.A. (2006). Epidemiology of Aphasia Attributable to First Ischemic Stroke: Incidence, Severity, Fluency, Etiology, and Thrombolysis. *Stroke, 37*, 1379-1384.

Gahl, S., Jurafsky, D., & Roland, D. (2004). Verb subcategorization frequencies: American English corpus data, methodological studies, and cross-corpus comparisons. *Behavior Research Methods, Instruments, & Computers 36*(3), 432-443.

Gahl, S. (2002). Lexical biases in aphasic sentence comprehension: An experimental and corpus linguistic study. *Aphasiology, 16*, 1173-1198.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences. *Journal of Memory and Language, 37*, 58-93.

Kim, M., & Thompson, C. K. (2000). Patterns of comprehension and production of nouns and verbs in agrammatism: Implications for lexical organization. *Brain and Language, 71*, 1-25.

MacWhinney, B., Fromm, D., Forbes, M. & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*, 1286-1307.

MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk. 3rd Edition.

Mahwah, NJ: Lawrence Erlbaum Associates.

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The Quantitative Analysis of

Agrammatic Production: Procedure and Data. *Brain and Language, 37*, 440-479.

Staub, A. (2007). The Parser Doesn't Ignore Intransitivity, After All. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 33(3), 550-569.

Webster, J., Franklin, S., and Howard, D. (2007). An analysis of thematic and phrasal

structure in people with aphasia: What more can we learn from the story of

Cinderella? *Journal of Neurolinguistics, 20*, 363-394.

Whitworth, A. (1995). Characterizing thematic role assignment in aphasic sentence

production – Procedures for elicited and spontaneous output. *European Journal*

of Disorders of Communication, 20(3), 384-399.

Table 1

Experimental Verb Tokens

Intransitively-based Verbs			Transitively-biased Verbs		
	Verb Frequency	Proportion Intransitive Use		Verb Frequency	Proportion Transitive Use
disappear	2.92	0.98	**determine**	2.57	0.35
grow	3.31	0.70	**teach**	3.33	0.34
dance	3.39	0.89	**clean**	3.54	0.47
fall	3.52	0.82	**lose**	3.63	0.63
bet	3.60	0.55	**buy**	3.63	0.83
stand	3.69	0.90	**read**	3.68	0.64
sit	3.75	0.97	**eat**	3.68	0.68
happen	3.75	0.93	**forget**	3.75	0.46
start	3.80	0.30	**hear**	3.86	0.51
talk	3.88	0.90	**find**	3.89	0.44
think	3.92	0.28	**want**	3.92	0.29

Verb frequency and proportion of (in)transitive use values came from Gahl et al. (2004).

Verb tokens were also matched for number of syllables.

Figure 1

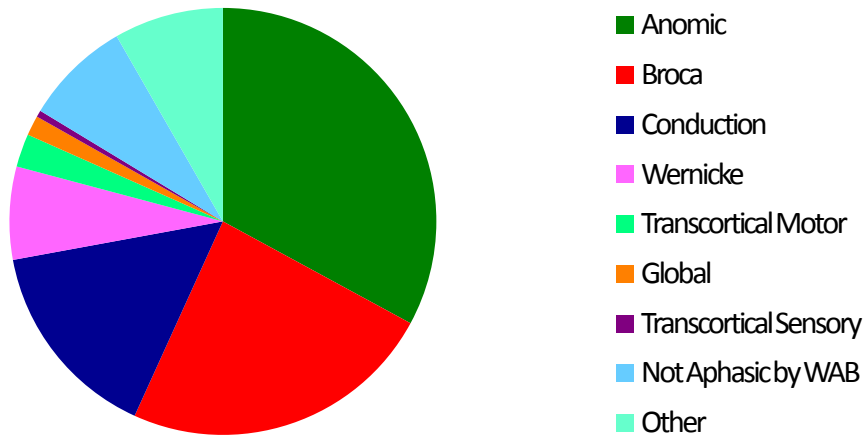Aphasia Types Represented in *AphasiaBank* (as of October 2014)

Figure 2

PWA's and Controls' Uses of Transitive Verbs in Various Sentential Structures

People with Aphasia   **Controls**
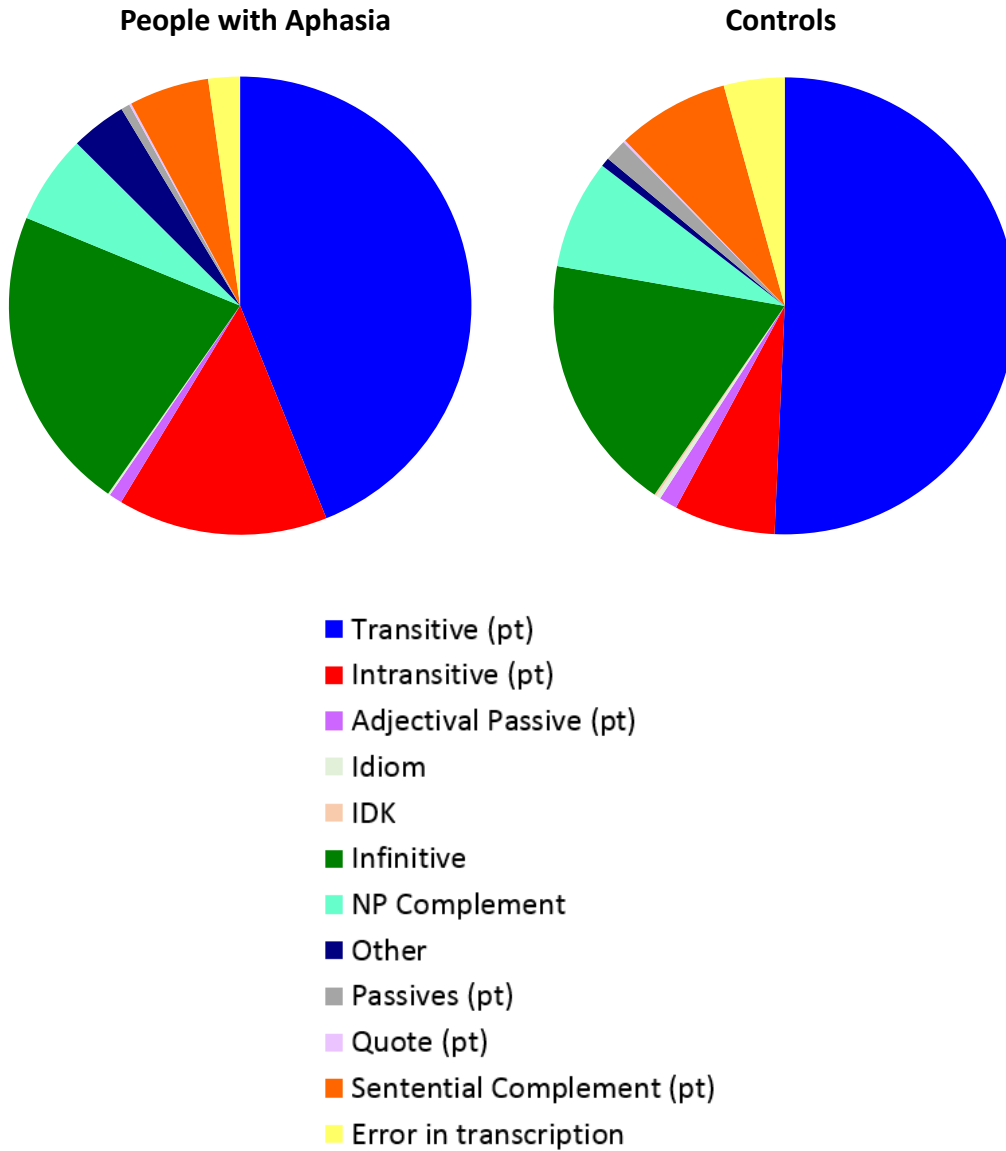

- ■ Transitive (pt)
- ■ Intransitive (pt)
- ■ Adjectival Passive (pt)
- ■ Idiom
- ■ IDK
- ■ Infinitive
- ■ NP Complement
- ■ Other
- ■ Passives (pt)
- ■ Quote (pt)
- ■ Sentential Complement (pt)
- ■ Error in transcription

Figure 3

PWA's and Controls' Uses of Intransitive Verbs in Various Sentential Structures

People with Aphasia   **Controls**

■ Transitive (pt)

■ Intransitive (pt)

■ Adjectival Passive (pt)

■ Idiom

■ IDK

■ Infinitive

■ NP Complement

■ Other

■ Passives (pt)

■ Quote (pt)

■ Sentential Complement (pt)
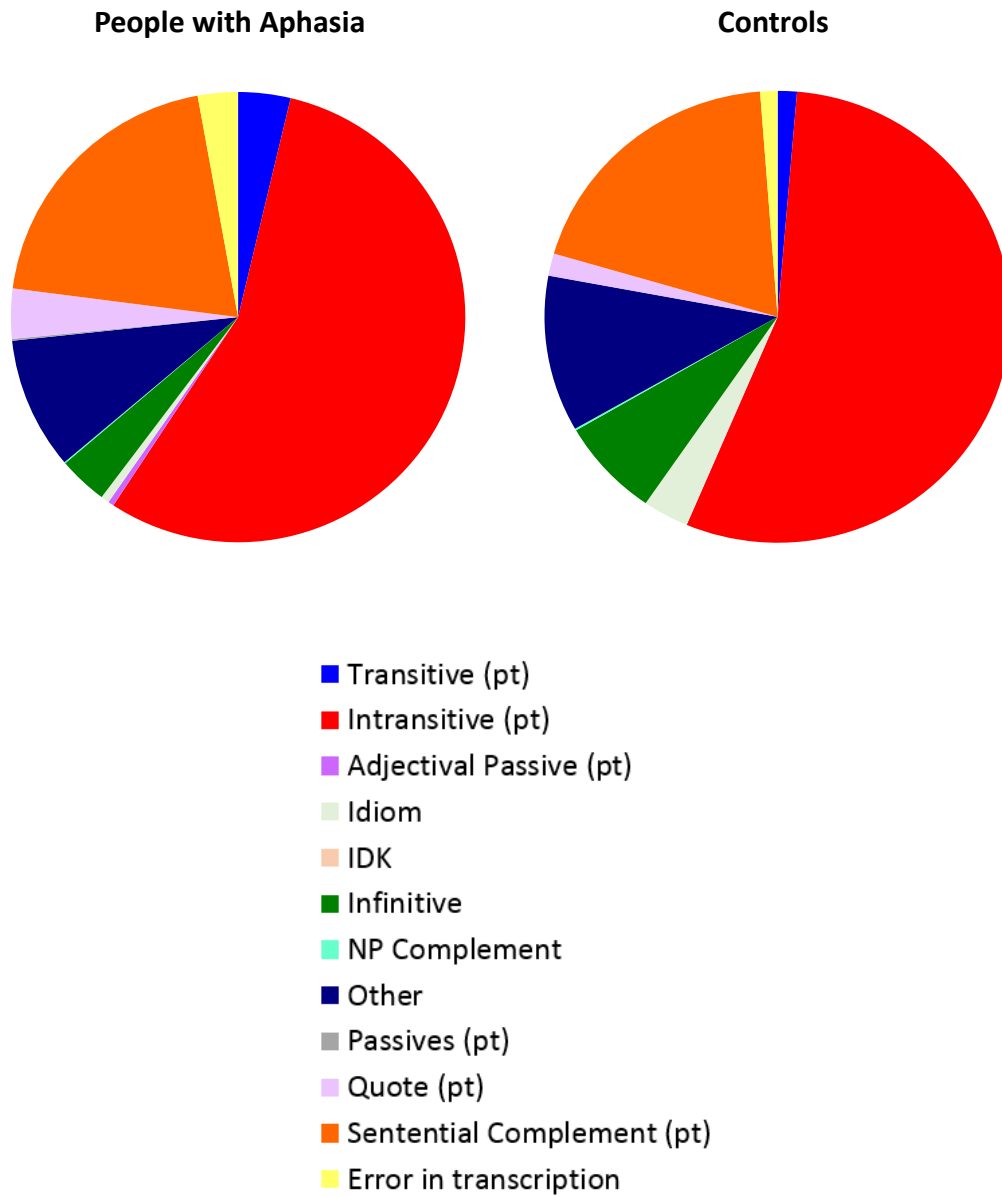
■ Error in transcription

Figure 4

PWA's and Control's Uses of Transitive and Intransitive Verbs (Regardless of Sentential
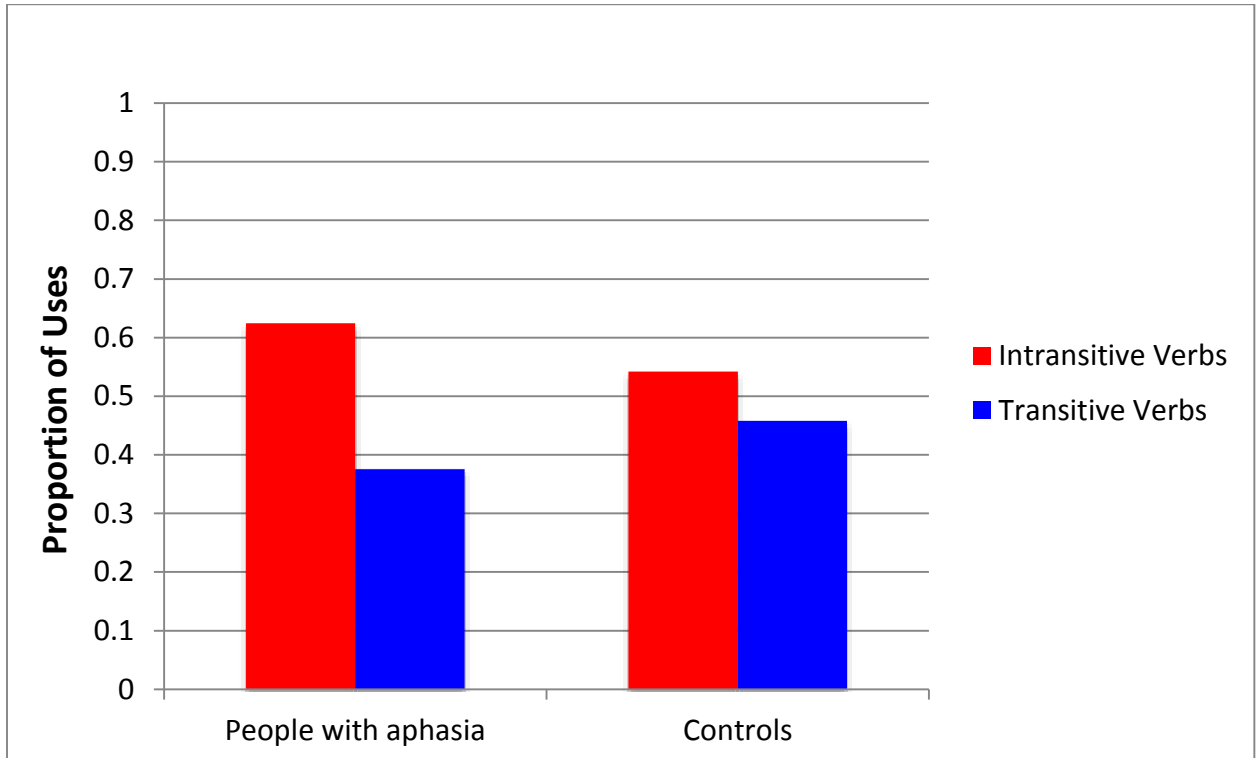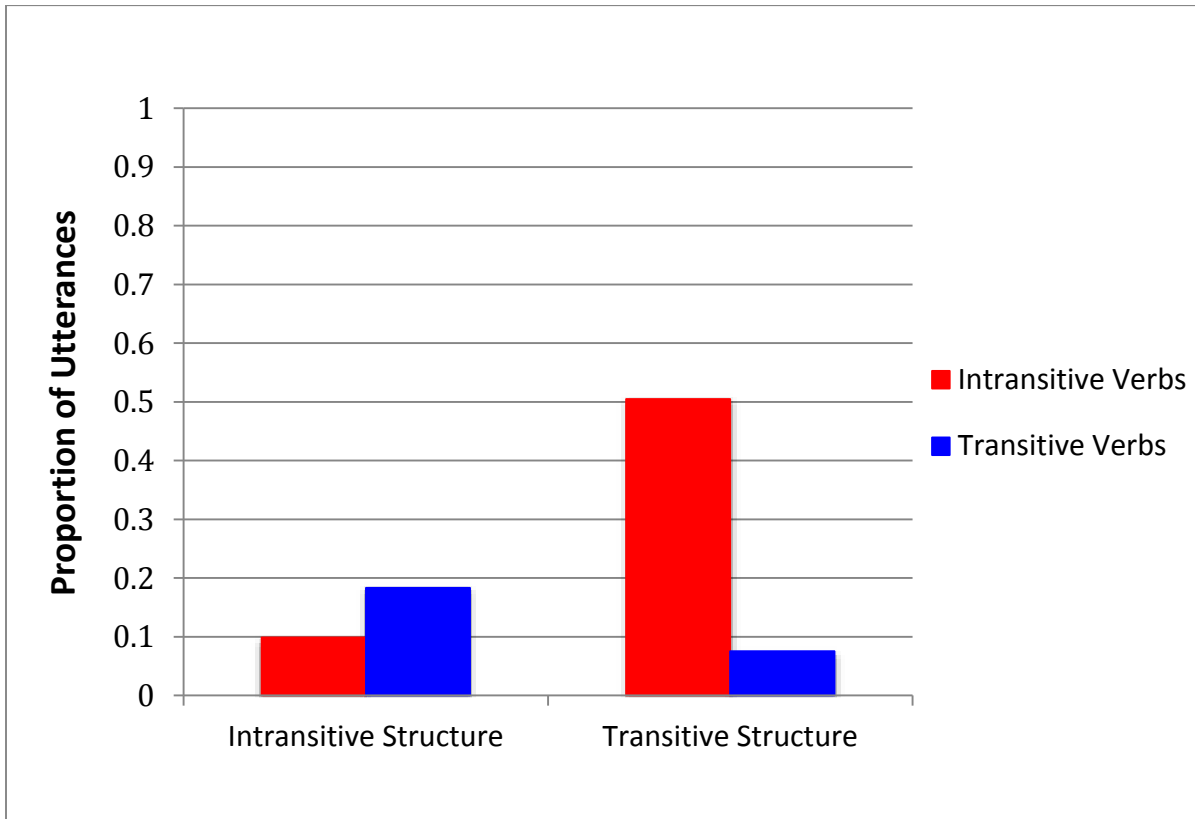
Structure)

Figure 5

Proportion of Utterances with Errors for PWA

Appendix A

Adapted Coding Scheme

Verb Codes

A: Verb Bias Codes (Without Error)

Under each utterance involving a target verb in the *PAR (participant) tier, we added a

tier, %gah, to insert verb bias codes. These codes followed the formula **verb=verb code**:

    EXAMPLE:

        PAR: *I <u>thought</u> I was saying, "Help!"*

        %gah: think=sfin.

           "Think" takes a sentential complement (sfin),

The verb codes on this tier largely followed the schema outlined by Gahl et al. (2004),

and consisted of:

Error Code	Description
tr	active transitive
trpt	active transitive with particle
in	active intransitive
inpt	active intransitive with particle
pass	passive
passpt	passive with particle
apass	adjectival passive
apasspt	adjectival passive with particle
sfin	sentential complement
sfinpt	sentential complement with particle
quote	quotation

quotept	quotation with particle
npcomp	noun phrase + sentential complement
npcomppt	noun phrase +sentential complement + particle
passcomp	passives with sentential complements of VP complements
inf	infinitive complements
other	Everything else
x	typos in transcription
i	idioms

The current study added a few considerations to the Gahl et al. (2004) coding scheme:

a. **"Other" code**

    a. We adapted the "Other" code to encompass ambiguous productions, since we could not assume the intended structure. For instance, we coded elliptical sentences as "Other," since participants don't actually produce the implicit phrase:

        ▪ PAR: *She would do anything they <u>wanted</u>*.

        %gah: want=other.

    b. We also considered some constructions to be non-communicative fillers, and marked them as "Other":

        ▪ PAR: *I (don't) <u>know</u>*.

        %gah: know=other.

        ▪ PAR: *I (don't) <u>remember</u>.*

        %gah: remember=other.

        ▪ PAR: *<u>See</u>, it was a long time ago.*

> %gah: see=other.

- PAR: *I talked with her for like nine days, <u>I think</u>.*

> %gah: think=other.

However, when these constructions (especially "know" and "remember") took on additional elements, they were coded according to the typical coding schema:

- PAR: *I <u>know </u>this*.

> %gah: know=tr.

- PAR: *I don't <u>remember</u> about it.*

> %gah: remember=in.

b. Surface level coding

Structures were primarily labeled according to what syntactic element immediately followed the verb. However, there were a few complications to this trend. We considered movement (e.g. Wh- movement) like Gahl et al. (2004), and also added considerations for:

a. *Adverbial phrases of time*

Noun phrases that followed a verb and conveyed a sense of time were considered adverbial phrases:

- PAR: *He <u>thought</u> all day about the problem.*

> %gah: think=in.

b. *Enough, much, and more*

Some phrases including "enough," "much," and "more" had to be interpreted according to the syntactic context.

- PAR: *The man is trying to get the cat but he doesn't know enough to get him.*

  %gah: know=in.

- PAR: *Now I'm probably going to tell you too* **much**...

  %gah: tell=tr.

- PAR: *So much short that need to want* **more** *gesture*

  %gah: want=tr.

c. *Indirect objects*

Verbs followed by indirect objects (and no direct object) were considered to be intransitive:

- PAR: *I asked you about the weather.*

  %gah: ask=in.

c. **Additional complementizers**

As a clarification, additional phrases (other than "that") were considered to be complementizers. The main focus was to identify and label sentential complements:

- PAR: *I don't know* **because** *he didn't say.*

  %gah: know=sfin.

- PAR: *I remember* **when** *it happened.*

  %gah: remember=sfin.

B: Verb Codes with Error

The current schema included considerations for how to interpret verb biases in the
presence of an error:

d. Abandoned Utterances

Abandoned utterances were defined as utterances where a structure was not
completed. These instances could either indicated by the original
AphasiaBank coder, when speech that followed the utterance was coded on
subsequent tiers, or within-tier abandonments, when a structure was recast
as a new structure within the utterance:

a. *Coder-indicated abandonment*

- PAR: *I don't <u>know</u> how…*

  %gah: know=other.

b. *Within-tier abandonment*

- PAR: *I <u>try</u> to [//] the something with my speech.*

  %gah: try=other.

e. Different omissions lead to different types of interpretation:

a. *Omissions of function words with one possible interpretation*

If a function word was omitted and there was only one possible
interpretation of that sentence, the verb was coded according to the
deeper layer:

- PAR: *I <u>want</u>… run.*

  %gah: want=inf.

The basic structure is infinitival even without the

preposition, "to."

b. *Omissions of content words with one possible interpretation*

If a content word was omitted and there was only one possible

interpretation of that sentence, the verb was coded according to the

surface layer:

- PAR: *I <u>want</u> to… the peanut butter.*

   %gah: want=other.

   Here, the infinitival "to" may be present, but we did not

   want to give credit for a verb that was not produced.

c. *Ambiguous omissions*

If a word was omitted and the sentence was structurally

uninterpretable, the verb was marked as an "Other" structure.

- PAR: *I <u>write</u> um the laptop and to my daughter every day.*

   %gah: write=other.

f. **Neologisms**

Along with AphasiaBank, we referred to non-words as "neologisms."

Neologisms were permissible as category fillers in the instance that there

was only one syntactic interpretation. Otherwise, utterances with these

phrases were marked as "Other". In most cases, we erred in favor of "Other":

- PAR: *I don't <u>want</u> him to mʌkəl onto a sandwich that's too big.*

   %gah: want=npcomp.

- PAR: _Seen, um dʌ._

    %gah: see=other.

Error Codes

In any ungrammatical production, we added a tier below the %gah tier to include error labels. Codes on this tier, %err, followed the formula **verb=verb code=error code**:

EXAMPLE:

PAR: _I think progress, but I don't know._

%gah: think=tr.

%err: think=tr=eav

"Think" occurs transitively, and there is a non-lexical error after the verb.

We developed the following error labels for this tier:

Error Code	Description
lbv	lexical error before verb
lav	lexical error after verb
omfb	omission of function word before verb
omfa	omission of function word after verb
minf	morphological and/or inflectional error
ebv	error before verb (non-lexical)
eav	error after verb (non-lexical)
pp	phonological paraphasia
sp	semantic paraphasia
aband	abandoned utterance
NR	not relevant

Let's clarify each error code with examples.

I.   **Lexical errors**

This category referred to unambiguously lexical errors – including semantic

and/or phonological issues – before or after the target verb. We coded lexical

errors for content words, but not function words:

- PAR: *I nont <u>know</u> how she got the… this…*

  %err: know=sfin=lbv.

- PAR: *because they don't <u>want</u> your faʊdɪs.*

  %err: want=tr=lav.

II.   **Omission of function word**

This category marked errors that were unambiguously the result of an omitted

function word, including auxiliary verbs:

- PAR: *Then the guy… you know, <u>dancing</u> and all of that…*

  %err: dance=in=omfb.

- PAR: *…and they <u>drove</u> her the party.*

  %err: drove=tr=omfa.

III. **Morphological and/or inflectional error**

These errors encompassed issues with number, agreement, and/or tense on the

target verb:

- PAR: *The cat is just <u>sit</u> there.*

  %err: sit=in=minf.

- PAR: *…and they <u>sawed</u> me at that time.*

%err: see=tr=minf.

- PAR: *…and so she <u>want</u> to get there really bad.*

  %err: want=inf=minf

IV.   Non-lexical errors

The next category (ebv and eav) served as a "catch all" for errors that were not

lexical, and did not clearly fit into the above categories. We continued to

consider deep layer representations of wh- movement in this category.

- PAR: *I went any time <u>tried</u>.*

  %err: try=in=ebv.

- PAR: *Is different now because my dad is <u>helping </u>me a much better*

  place.

  %err: help=tr=eav.

- PAR: *I <u>think </u>is talking.*

  %err: talk=sfin=eav.

- PAR: *Some what I just <u>told</u> you.*

  %err: tell=tr=eav.

V.   Phonological paraphasia

These errors included phonological distortions of the target verb.

- PAR: *I <u>sɔt</u> something was wrong with my arm.*

  %err: think=sfin=pp.

- PAR: and *I <u>pɛl</u> asleep.*

  %err: fall=in=pp.

VI.   Semantic paraphasia

These included semantic substitutions in which the produced, target verb did not

make sense within the structure. However, there were no instances in which we

used this code throughout the database. This may be an example of a label that

we revise, as it may be that a few of our existing codes already capture this

error.

VII.   Abandoned Utterance

Related to the discussion of abandoned utterances above, we created a separate

category for abandoned utterances so as not to assume the intended structure.

- PAR: *He doesn't <u>realize</u> they….*

    %err: realize=other=aband.

- PAR: *I don't <u>know</u> what made Cinderella…*

    %err: know=sfin=aband.

VIII.   Not relevant

This code allowed us to exclude errors that did not affect the target verb phrase,

and productions that were recast grammatically. However, since we added this

code in the error coding phase, it may be that some recasts were marked as

error and others not. We can reconsider how to code recasts beginning at an

earlier stage in coding.

- PAR: *Mother was dead but someone <u>thought</u> it and was in…*

    %err: think=tr=nr.

One final consideration is that, more often than not, an ungrammatical production could have more than one error. We labeled every possible error interpretation, such that many utterances had multiple error codes:

- PAR: *Kicking a soccer.*

  %err: kick=tr=lav=ebv.

- PAR: *and little mouses is helping decorate… dress.*

  %err: help=inf=lbv=minf=omfa.

- PAR: *…and believing that the second people should…*

  %err: believe=sfin=ebv=omfb=eav=aband.

In our final calculations, we decided to collapse across error types. We separated uses of a target verb with error from those without error. So, for instance, each of the last three error examples would count simply as one target verb production with error. As mentioned in the Limitations, it would be informative to use these finer grain error codes to characterize PWA's speech – and we will continue revising them as we better understand PWA's grammatical profile.