Title: Investigating the origin of nonfluency in aphasia:

A path modeling approach to neuropsychology

Authors: Nazbanou (first name) Nozari (last name)[1,2] , Yasmeen (first name) Faroqi-Shah (last name)[3]

Affiliations:

[1]Department of Neurology, Johns Hopkins School of Medicine; [2]Department of Cognitive Science, Johns Hopkins University (1629 Thames Street, Suite 350, Baltimore, MD 2123, USA; Phone: 443-287-1712; Fax: 410-955-0188; Email: nozari@jhu.edu)

[3]Department of Hearing and Speech Sciences, University of Maryland (0141 F, Lefrak Hall, 7251 Preinkert Dr., College Park, MD 20742, USA; Phone: 301- 405-4229; Email: yfshah@umd.edu)

Corresponding author: Nazbanou Nozari (nozari@jhu.edu; nazbanou.nozari@gmail.com)

Abstract

A major challenge in understanding the origin of clinical symptoms in neuropsychological impairments is capturing the complexity of the underlying cognitive structure. This paper presents a practical guide to path modeling, a statistical approach that is well-suited for modeling multivariate outcomes with a multi-factorial origin. We discuss a step-by-step application of such a model to the problem of nonfluency in aphasia. Individuals with aphasia are often classified into fluent and nonfluent groups for both clinical and research purposes, but despite a large body of research on the topic, the origin of nonfluency remains obscure. We propose a model of nonfluency inspired by the psycholinguistic approach to sentence production, review several bodies of work that have independently suggested a relationship between fluency and various elements in this model, and implement it using path modeling on data from 112 individuals with aphasia from the AphasiaBank. The results show that word production, comprehension, and working memory deficits all contribute to nonfluency, in addition to syntactic impairment which has a strong and direct impact on fluency. More generally, we demonstrate that a path model is an excellent tool for exploring complex neuropsychological symptoms such as nonfluency.

Keywords: nonfluent aphasia, speech fluency, agrammatism, AphasiaBank, path model, structural equation model

1. Introduction

1.1. The problem of understanding the origin of multi-factorial effects

As goal-oriented agents, we are often interested in achieving a particular outcome. Many outcomes, however, result from complex interactions between several factors, and this creates difficulty in tracing an outcome back to its origin(s). The same difficulty applies to understanding the origins of neuropsychological disorders, and the problem is compounded by the fact that the outcome measure is not always so easy to define. Many disorders, manifest as a combination of various symptoms with different degrees of severity among individuals. Thus, not only do we face a complex network of underlying factors as predictors, but also a heterogeneous outcome variable that cannot be reduced to a simple binary measure. This paper discusses a statistical method of analysis, *partial least square - path modeling (pls-pm)* that allows for simultaneous representation of a complex network of predictors with various interdependencies among them, predicting a multivariate dependent variable that can itself be represented by more than one outcome measure.

Rather than a mathematical introduction to path modeling, which exists elsewhere (Lohmöller, 2013), this paper aims to present a practical guide for the application of this method to neuropsychological data. To this end, we use the example of nonfluency in individuals with aphasia, and use path modeling to investigate the origin(s) of this heterogeneous problem. In the next section, we first review the literature on the potential contribution of various factors to nonfluency in aphasia to make a clear case for why a comprehensive study of the origins of nonfluency can benefit from pls-pm. We then present a general overview of pls-pm, followed by the application of the technique to a dataset of 112 individuals with aphasia from the AphasiaBank (MacWhinney, Fromm, Forbes, & Holland, 2011) and discuss the steps involved in model building, model evaluation, and ultimately interpretation of the results.

1.2. Nonfluency in aphasia

Persons with aphasia (PWA) exhibit a variety of language symptoms, such as impaired word retrieval, auditory processing, comprehension, and sentence formulation. The presence and severity of these symptoms vary between individuals and is the basis of classifying aphasias into clinical subtypes (e.g., Broca's, conduction, Wernicke's, anomic, etc.). One language symptom, nonfluency, has assumed a central role in aphasia classification following the work of Benson (1967). In a spontaneous speech analysis of 100 PWA, Benson (1967) examined 10 different symptoms, such as word choice, rate of speaking, articulation, phrase length and effort (initially used by Howes & Geschwind, 1964). He found a bimodal clustering of symptoms, such that individuals with left anterior lesions had slow rate of speech, spoke effortfully and predominantly used substantive words, while persons with left posterior lesions spoke effortlessly, but lacked substantive words and instead produced paraphasias. The bimodal distribution based on spontaneous speech symptoms was later replicated using more conservative fluency criteria and statistical analyses (Kerschensteiner, Poeck, & Brunner, 1972). Thus, the speech nonfluency-fluency dichotomy became a hallmark of the different syndromes arising from anterior versus posterior lesions in PWA.

Rate of speech, typically expressed as the number of uninterrupted words spoken per minute (words per minute or wpm), is an important criterion for classifying persons as fluent or nonfluent. Based on a normal speech rate of 100 to 175 wpm, the cut-off for nonfluent PWA was set as less than 50 (Benson, 1967) or 100 (Howes, 1964) wpm. These authors also proposed a hyperfluent group with rate of speech greater than 150 (Benson, 1967) or 175 (Howes, 1964) wpm. Kerchensteiner et al.'s (1972) analysis questioned the existence of hyperfluent PWA and they proposed very slow (<50 wpm), slow (50-90 wpm) and normal rates of speech. While wpm is an objective measure, it is not necessarily the best index of nonfluency because perceived fluency in communication depends not only on how fast the

words are produced, but also on the length, complexity and grammaticality of phrases, presence of paraphasic and stereotypical utterances, and prosodic abnormalities (Benson, 1967; Feyereisen, Pillon, de Partz, 1991; Kerchensteiner et al., 1972. Indeed, standard aphasia test batteries use a multidimensional rating system that takes several of these variables into account in order to assess the degree of fluency in PWA (Goodglass & Kaplan, 1983; Kertesz, 1982). The *Boston Diagnostic Aphasia Examination* (BDAE) uses a 7-point Likert scale for each of six fluency characteristics (Goodglass & Kaplan, 1983; Goodglass, Kaplan, & Barressi, 2001). The *Western Aphasia Battery* (WAB) uses a single 10-point scale in which each point represents a cluster of characteristics (Kertesz, 1982, 2006; see Appendix A). Agreement between the fluency classifications on WAB and BDAE is about 84% (Wertz, Deal, & Robinson, 1984), most likely because of the different ways in which fluency is rated in these test batteries.

However, despite its canonical role in clinical classification and existence of several standardized tests for its assessment, the origin of nonfluency in aphasia has been difficult to pin down. This is partly because of the different weights given by examiners to different aspects of the problem. For example, some researchers have focused on the rate of speech, defining fluent speech as composed of many words produced over a long time period with a rapid rate (Kreindler, Mihailescu, & Fradis, 1980). Others have focused on both rate and syntactic complexity of the narrative (Goodglass & Kaplan, 1983; Wagenaar, Snow, & Prins, 1975). Still others have emphasized semantic richness, such as the ability to speak in a coherent, creative, semantically dense, and contextually appropriate manner (Fillmore, 1979). In keeping with these diverse definitions, several studies have documented the notorious inconsistency with which listeners make fluency judgments of PWA (Gordon, 1998; Holland, Fromm & Swindell, 1986; Trupe, 1984), as well as the absence of high inter-rater reliability between aphasia experts in rating fluency (Holland, Fromm, and Swindell, 1986; Trupe, 1984). In a regression analysis, Park et al. (2011) showed that rating judgments were influenced by speech rate, speech productivity (amount of time

spent speaking), and audible struggle (degree of vocal tension and articulatory effort perceived in the audio recordings), marking the multi-faceted nature of nonfluency from the listener's perspective.

There is an important message in these discrepancies: speech fluency is most likely a multi-faceted problem with more than one origin. As such, gaining a holistic understanding of the problem of nonfluency requires a model that can accommodate several factors and their relationships in giving rise to fluent speech. In the next section, we use a psycholinguistic framework to determine the likely contributors to nonfluency that should be included in such a model.

1.3. Deconstructing Fluency

Language production models identify several processing stages that must be completed to produce fluent speech (Bock, 1995; Garrett, 1975; Griffin & Ferreira, 2006; Levelt, 1999). Following the decision of what to express (conceptualization), language production involves selection of lexical items (a process that is informed by the syntactic properties of the to-be-retrieved word), formulation of the utterance type and structure (question or statement, active or passive sentence, etc.), phonological encoding, articulatory planning, and self-monitoring (Levelt, 1989; Nozari, Dell, & Schwartz, 2011; Nozari & Novick, in press). Thus, a glitch in any of these processes could theoretically disrupt the fluency of speech. However, relatively little work has examined the contribution of these psycholinguistic processes to (non)fluency in aphasia, especially given the intriguing finding that overall aphasia severity does not predict fluency (Gordon, 2006). In the following paragraphs, we briefly review variables that are likely to contribute to fluency in PWA, supplemented by the discussion of relevant research from other fields with a focus on nonfluency, such as stuttering. Importantly, the scope of the current study is limited to the exploration of cognitive and linguistic factors underlying nonfluency. For this reason, as well as the limitations of the data bank used for the analyses reported in this paper (see below), we do not investigate the influence of motor problems or speech apraxia on nonfluency. This does not imply

that we view such factors as inconsequential to nonfluency, but rather as operating at levels outside of the scope of the current study. Indeed, we argue in the Discussion section of the paper that speech apraxia is undoubtedly relevant to nonfluency and its contribution requires further investigation.

Word production. Since words are units of a sentence, it is plausible to hypothesize that problems with producing single words can affect fluency of speech. Word production requires at least two steps (Dell, 1986; Garrett, 1975; see Dell, Nozari & Oppenheim, 2014 for a review): lexical retrieval and phonological encoding (and later articulation), and PWA can be impaired in either or both stages (e.g., Schwartz, Dell, Martin, Gahl, & Sobel, 2006). Several studies point to the association between impaired word production and nonfluency. In neurotypical individuals, lexical access problems lead to more disfluencies (e.g., Hartsuiker & Notebaert, 2009). Disrupted lexical retrieval has also been implicated in disfluencies associated with stuttering (Postma & Kolk, 1993; Prins, Main, & Wampler, 1997; Wingate, 1988). Some studies have found slower verb retrieval (Prins et al., 1997) and slower covert noun retrieval (Sasisekaran, De Nil, Smyth & Johnson, 2006) in persons with stuttering compared to healthy controls. But this finding is not always replicated (e.g., Hennessey, Nang & Beilby, 2008; Tsai & Ratner, 2016). Thus the relationship between word retrieval and fluency warrants further investigation.

In aphasia, Wilshire et al. (2000) proposed a relationship between control deficits in single word production and nonfluent aphasia. Indirect evidence for the association between picture naming and speech rate (wpm) comes from volumetric lesion analyses, in which both picture naming (assessed with the Boston Naming Test, BNT, Goodglass, Kaplan & Barresi, 2001) and speech rate were associated with left arcuate fasciculus lesion volume (Marchina et al., 2011; Wang, Marchina, Norton, Wan & Schlaug, 2013). But although association with the same lesion may imply a co-dependence, these studies did not report the direct correlation, or a causal relation, between BNT performance and speech rate. On the behavioral side, several studies have documented impaired action (verb) naming in agrammatic aphasia,

which is often accompanied by nonfluency (e.g., Gordon, 2006; McCarthy & Warrington, 1985; Luzzatti et al., 2002; Zingeser & Berndt, 1990), however, this finding is not always replicated. For example, Bastiaanse and Jonkers (1998) found no difference between agrammatic and anomic PWA in their production of either object or action names in isolation. A possible relationship also exists between problems of phonological encoding and fluency: Andreetta and Marini (2015) found that the production of phonological errors (mainly false starts) was negatively correlated the proportion of complete sentences. As noted before, grammatical completeness is one criterion used to differentiate nonfluent from fluent aphasia (Goodglass & Kaplan, 1983; Kertesz, 1982). Thus phonological difficulties could affect grammatical completeness, which in turn could affect (objectively measured) speech rate and the subjective judgment of fluency. Although Andreetta and Marini (2015) did not report correlations between phonological errors and speech rate, their PWA group had slowed speech rate (mean of 57.6 wpm).

To summarize, it is theoretically plausible that problems with single word production cause nonfluency in aphasia, although the evidence for a direct relationship between the two is mixed. Moreover, most of these studies do not simultaneously account for other factors that could affect fluency. In the current path modeling analysis, we investigate this relationship in a model that also contains other factors which could potentially contribute to nonfluency. While AphasiaBank (MacWhinney et al., 2011) does not contain measures that allow us to separate lexical retrieval from phonological encoding, we were able to use three confrontation naming tests typically used in aphasia to represent the latent variable of word production: the BNT, object naming of the WAB and Verb Naming Test (VNT, Cho-Reyes & Thompson, 2012). Inclusion of both object and verb naming in this category is justified by the evidence suggesting that grammatical differences between nouns and verbs are not pertinent to picture naming in isolation (e.g., e.g., Pechmann et al., 2004; Vigliocco et al., 2005;

see Vigliocco, Vinson, Druks, Barber & Cappa, 2011 for a review). We will revisit the appropriateness of this choice when applying the model checks.

Syntactic production. Planning sentences and multi-word utterances not only places higher computational demands than single word production, but also relies to a greater extent on timely execution of preceding processing stages and maintaining inter-dependencies between different parts of speech (Bock, 1995). Children, adolescents and adults are known to produce more disfluencies on syntactically complex compared to simple sentences (e.g., Haynes & Hood, 1978; Ratner & Sih, 1987; Silverman & Ratner, 1997; Spieler & Griffin, 2006), and disfluent utterances are more complex and longer than fluent utterances in children who stutter (Melnick & Conture, 2000).

Clinical observation of individuals with Broca's aphasia has generated an intuition about the association between agrammatism and nonfluency. However, the link between disfluency (or rather nonfluency) and syntactic production in aphasia is not so well-established. Several studies have demonstrated morphosyntactic impairments in nonfluent aphasia (e.g., Faroqi-Shah & Thompson, 2003; Goodglass, Christiansen & Gallagher, 1993; Kolk & Heeschen, 1996; Saffran, Berndt & Schwartz, 1989), but grammatical problems have also been reported in fluent aphasia (Bird & Franklin, 1996; Edwards, 1995; 1998; Edwards & Bastiaanse, 1998; Edwards, Bastiaanse, & Kiss, 1994; Faroqi-Shah & Thompson, 2003), questioning the close correspondence between agrammatism and nonfluency. Moreover, these studies use a dichotomous classification of aphasia, and it is unclear whether the magnitude of morphosyntactic impairment is correlated with the severity of nonfluency.

Although the concept of syntactic complexity varies across researchers and is hotly debated (Pallotti, 2015), it typically refers to the number of matrix clauses used to derive a sentence, or the structural complexity of the verb phrase, or the number/types of morphosyntactic elements used. Some

views of syntactic complexity consider developmental sequences, for example they view 3rd person singular *-s* to be more complex than progressive *-ing* in English because the former is acquired later (Brown, 1973). In a recent study, Thorne and Faroqi-Shah (2016) found that syntactic complexity measures that weigh developmental stages are sensitive predictors of syntactic abilities in aphasia. In other cases, syntactic complexity may just refer to the length of the sentence. Thus, the metrics of syntactic complexity vary with the theoretical orientation of the researcher, as well as the population under study. Given that syntactic complexity is not a unitary concept, we chose a variety of measures in this study to include counts of grammatical utterances (%grammatical), grammatical complexity that weighs developmental stages (Developmental Sentence Score, DSS, Lee, & Canter, 1971), presence of morphological elements (bound and free morphemes represented by % past tense and % prepositions respectively), and the proportion of verbs (%verb).

Comprehension. At its core, language comprehension reflects the integrity of (and access to) semantic representations. Given that fluent language production entails efficient access to (robust) semantic representations, one could predict a link between language comprehension and fluent speech. Although the receptive-expressive dichotomy of aphasia was replaced by the fluent-nonfluent dichotomy in the mid-20th century (see Benson, 1967), comprehension and fluency are dissociable in aphasia. PWA could have low (e.g., Wernicke's aphasia) or relatively spared (e.g., conduction or anomic aphasia) language comprehension, and nonfluent aphasia itself could be associated with low (e.g. global aphasia) or relatively spared (e.g., Broca's aphasia) comprehension (Martin, 1987). Although, people with agrammatic aphasia demonstrate at least some problems with sentence comprehension (e.g., Schwartz, Saffran, & Marin, 1980), a direct relationship between such problems and fluency has not been tested. In the present path modeling analysis, we examine this relationship using two measures from WAB that index word and sentence comprehension.

Working memory. The ability to maintain and manipulate linguistic units (i.e., verbal working memory) is crucial for language comprehension (e.g., Caplan & Waters, 1999; Just & Carpenter, 1992) and production (Acheson & MacDonald, 2009; Ellis, 1980). Given the multi-step nature of sentence processing, verbal working memory helps maintain the connections between multiple processing levels. This is particularly important for syntactic production, which requires accurate representation of dependencies between different parts of the sentence (e.g., singular subject → add singular morpheme to the verb; event in the past → retrieve the past form for the verb, etc.). Therefore, working memory limitations could in theory cause linguistic representations to decay before becoming available for later planning/processing, and thus impact speech fluency.

In PWA, numerous researchers have reported limited working memory capacity (e.g., Caspari, Parkinson, LaPointe, & Katz, 1998; Ivanova, Dragoy, Kuptsova, Ulicheva & Laurinavichyute, 2015; Martin, Shelton, & Yaffee, 1994; Tompkins, Bloise, Timko & Baumgaertner, 1994). One study found that the relationship between working memory and comprehension was strong only in nonfluent aphasia, but not in fluent aphasia, suggesting a three-way interaction between comprehension, working memory and fluency (Ivanova et al., 2015). As is the case with most PWA research reviewed in this paper, fluency was not measured, but was inferred from aphasia subtype. However, these findings illustrate the need for further research into the role of working memory in producing fluent speech in aphasia. In our model, we include several measures from the AphasiaBank (MacWhinney et al., 2011) that act as proxies for working memory abilities.

Self-monitoring and repair. Self-monitoring refers to the process of checking one's own production in order to identify errors, initiate repairs and prevent errors from occurring (Nozari & Novick, in press; Sampson & Faroqi-Shah, 2011). Detection of an error can disrupt the normal process of production by signaling the need to halt ongoing production and start the process of repair (Hartsuiker & Kolk, 2001). A recent study in PWA has reported that blocking self-monitoring through the auditory channel could

increase fluency (Jacks & Haley, 2015). It is thus possible that nonfluency is at least in part due to monitoring problems. Ideally, monitoring is represented by the proportion of detected errors over a large number of trials (e.g., Nozari et al., 2011). Unfortunately the database we used did not allow us to derive this measure. Instead, we were able to include repetitions (when a certain word was repeated immediately) and retraces (when a phrase was repeated often with some modification from the original utterance which may or may not have been incorrect) during the narration of the Cinderella story. It must be noted, however, than repetition and retraces do not necessarily represent monitoring abilities, but could instead indicate problems in planning or uttering the rest of the sentence. Thus in later sections we refer to this variable as “restart” instead of “monitoring”.

To summarize, a vast body of literature underscores the central role of fluency in characterizing aphasia. Research from different populations, suggests a possible role for several factors including comprehension, lexical retrieval, phonological encoding, syntactic processing, working memory, and self-monitoring in shaping fluent speech. While each of these processes has been investigated individually in aphasia, their direct influence on fluency as well as the interactions among them in giving rise to fluent speech is yet to be thoroughly examined. This study aims to do that.

2. A path modeling approach

2.1. Pls-pm vs. regression

In a nutshell, pls-pm is a partial least square approach to Structural Equation Modeling (SEM; Kline, 2015). It is similar to regression in spirit, in that a number of independent variables (predictors) are tested for their predictive power over a dependent variable (outcome), except that the construct of both the predictors and the outcome could be more complex in pls-pm. Two differences are noteworthy between simple regression and pls-pm models (1) the distinction between *latent* and *manifest* variables

in pls-pm, and (2) the possibility of interdependence of the predictors in pls-pm. We unpack these two differences below.

Latent vs. manifest variables. The predictor and outcome variables are often compound constructs called “latent variables” each represented by several observed scores and measures called “manifest variables”. This distinction is theoretically important. Latent variables are often abstract constructs, such as working memory or nonfluency; they cannot be directly observed or measured. Instead, they can be indirectly assessed through tasks that are thought to tap into such abilities or impairments. Manifest variables, on the other hand, are direct measurements such as scores on a cognitive test, but they often do not represent a pure ability or impairment because they depend, to some degree, on the other cognitive processes involved in the tasks in which they are measured. This so-called task-impurity problem (Burgess, 1997; Phillips, 1997) precludes the use of measurements from a single task as a pure proxy for a latent variable of interest.

One way to get around the task impurity problem is to represent the latent variable of interest (e.g., working memory) with more than one manifest variable measured in tasks that all tap into the same latent variable but involve different basic processes (e.g., visual vs. auditory working memory tasks). This way, when estimating the latent variable, the portion of variance in performance that truly reflects the latent variable of interest is extracted from the manifest variables, overshadowing variance due to task-specific processes. But why not use a linear regression method and enter all the manifest variables in the model? Note that if all the manifest variables used to represent a latent variable are indeed measuring the same variable, their scores would be highly correlated with one another. This is known as the problem of multicollinearity in regression (Farrar & Glauber, 1967), which leads to unstable estimates for the coefficients. The only solution to multicollinarity in regression models is to generate composite variables through some form of averaging of scores from the different manifest variables, but such averaging will inevitably lead to loss of invaluable information. The pls-pm

framework solves this problem by defining a hierarchical structure for the latent and manifest variables, such that the common variance of all the manifest variables representing a certain latent variable is represented by that single variable, however that latent variable will have a much more accurate estimate because the task-impurity problem has been minimized.

Interdependence of predictors. The earlier review of the literature suggested several factors as potential predictors of nonfluency. But a closer look at some of these variables suggests that they might not be independent of one another. For example, it will not be surprising to find out that individuals with poor comprehension have difficulty in picture naming, perhaps because they have lost the semantic knowledge of the object to be named, or they may not understand the task instructions. Similarly, individuals with poor working memory may be those who have trouble with certain aspects of syntactic production. Working memory demands are shown to be particularly high for syntactically complex sentences (e.g., passives and object relatives) because of the non-transparent mapping between surface structure and underlying conceptual representations (e.g., agent→action→theme; Just & Carpenter, 1992; Sung et al., 2009), or the need to keep track of dependencies such as subject-verb agreement (Hartsuiker & Barkhuysen, 2006).

Finally, depending on how a latent variable is measured, it may show dependencies on other variables. For example, working memory measured as recall of words and sentences is likely to depend on both comprehension and word production abilities. While a simple regression model tests the dependence of the outcome variable on the predictors, it cannot accommodate interdependencies between predictors. One option would be to ignore such dependencies altogether. The obvious problem with this approach is that the model is not a realistic one. A second option is to run a series of independent regression models, each having one of the predictors as its dependent variable, and use the residuals of these models in the model with the outcome variable as its dependent variable. The most obvious problem with this approach is that it is very time-consuming. Pls-pm allows

interdependencies between predictors in the network, by simultaneously running a series of multiple regressions and estimating the parameters for these models in parallel. In other words, the deviation between the observed and the estimated pattern is not only minimized at the level of local models, but globally for all models within the network. In summary, pls-pm is a statistical method for analyzing the relationship between latent and manifest variables as a network of multiple interconnected linear regressions.

2.2. The inner model vs. the outer model

A path model consists of two sub-models: the *inner model* and the *outer model*. The inner model characterizes the relationship between the latent variables. This is the part of the model that tests the hypotheses regarding the effect of the independent variables on the outcome variable of interest. For the nonfluency model, the inner model consists of nonfluency as the dependent variable, and word production ability, syntactic production ability, comprehension, working memory and restarts as independent variables. As can be seen in Figure 1, it is possible to define dependencies between independent variables, as well as between each independent variable and the outcome variable.

The outer model reflects the relationship between the latent variables (conventionally shown as circles) and their manifest variables (conventionally shown as rectangles). The direction of the arrows between the latent and manifest variables depends on the theoretical assumptions behind the relationship between the two. In some cases, latent variables cause the manifest variables. For example, working memory ability is the reason that a person obtains a certain score on working memory tasks. This is called the *reflective* mode. It is, however, possible to define a latent variable that is formed by several manifest variables. For example, one can model the risk of a heart attack as a latent variable formed by manifest variables such as number of cigarettes smoked per day, blood pressure level,

cholesterol level, etc. In this case, the risk is not causing the measurements on the manifest variables, but is being formed by them. This is called the *formative* mode. Most cognitive models, including the example in this paper, contain reflective latent variables.

--- Insert Figure 1 about here ---

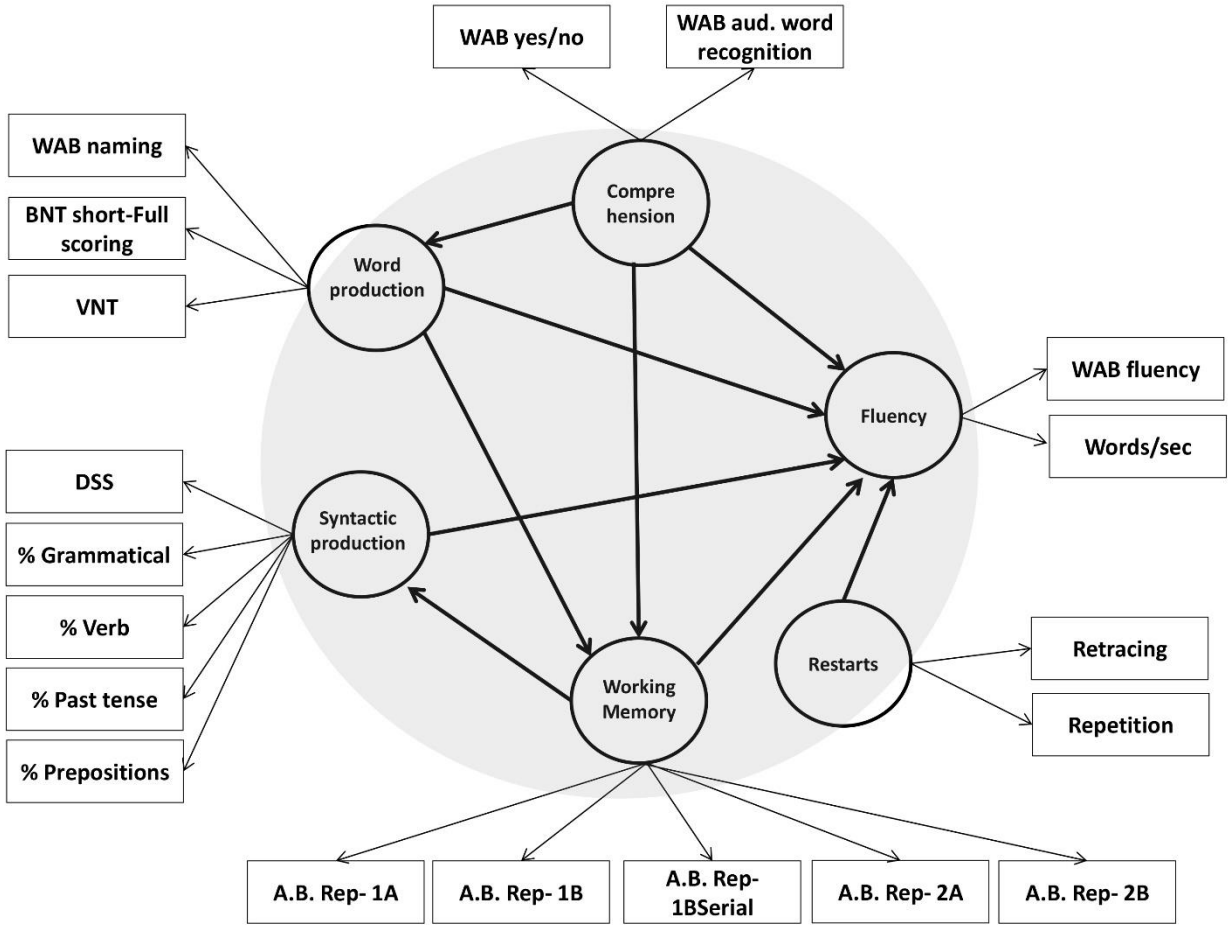

Figure 1- The architecture of the fluency pls-pm model. The circles show latent variables, and the rectangles the manifest variables (see also Table 2). The large gray circle represents the inner model. The direction of the arrows indicates which variable exerts an influence over the other variables. Note that the latent variables in the model are all reflective.

3. Using pls-pm to understand the origin of nonfluency in aphasia

There are several software packages available for conducting pls-pm. All the analyses reported in this paper were carried out using the free software R, with the plspm() package (Sanchez, 2013).

3.1. Dataset, variables, and model architecture

To generate reliable estimates, path models require a large dataset. We thus used the AphasiaBank database which contains data from individuals with various types and severities of aphasia who have been tested and scored using a standard protocol (see MacWhinney et al., 2011, for details). While the tests included in the battery are not always ideal for representing the variables included in our model, the homogeneity of the tests and the scoring systems together with an unusually large sample size for a neuropsychological study makes this database a great first step for studying the origins of nonfluency. The current study included 112 English-speaking individuals with post-stroke aphasia (48% women, Mean age = 60.3, SD = 13.1; Mean years of education = 15.8, SD = 2.8; Average number of years post-stroke = 5.49, SD = 4.02), all native speakers of American or Canadian English (88% monolingual), who had all or most of the measures required for our manifest variables (see below).

3.1.1. Measuring nonfluency

Nonfluency in the database was defined by the WAB-R spontaneous speech fluency score (Kertesz, 2006). The WAB-R spontaneous speech sample is obtained from PWA's answers to six

questions and description of a picnic scene, and fluency is rated on a 10-point scale (Appendix A). Glancing at the scoring system it is evident that (a) fluency scores are somewhat subjective and dependent on the examiner's judgment, and (b) fluency scores are influenced by paraphasias and grammatical competence. To address (a), before entering the WAB-R fluency scores into the model, we checked whether non-cognitive factors, such as age, gender, or educational level of the individual with aphasia or the examiners' level of experience may have biased the scores. Table 1 shows the results of this analysis. In a simple regression model, participants' age, gender, level of education, or years post-stroke did not reliably predict their fluency scores. Neither did the examiners' years of experience working with individuals with aphasia. The only factor that significantly affected scores was the years that the individual had received speech therapy. This analysis reassured us that the WAB-R fluency scores were not biased by factors irrelevant to the participants' language impairment.

In the earlier section, we discussed the rationale for using a nonfluency measure that represents several aspects of the problem, and WAB-R fluency score is such a measure. Nevertheless, to decrease subjectivity and to address the influence of paraphasia and grammatical problems on nonfluency, we complemented the WAB-R fluency scores with a secondary measure of fluency, spoken words/second, which we calculated using the clips and scripts from retelling of the Cinderella story (part of the AphasiaBank protocol). Videos of PWA's Cinderella story retelling were viewed in ELAN (http://tla.mpi.nl/tools/tla-tools/elan/; Sloetjes & Wittenburg, 2008), and the total time on task was registered for each individual after removing examiner's interjections. The total number of words spoken during this time was obtained from EVAL, which is a CLAN analysis tool (MacWhinney, 2000). The number of words divided by the time spent on task was used as our words/second measure of fluency. Note that this measure, unlike WAB-R spontaneous fluency score, does not give special weight to paraphasias or agrammatic speech. It simply counts the number of unrepeated words spoken in a finite time window. Figure 2 shows the box plots for the two fluency measures. Fluency measured with

WAB-R scores had a mean of 7.39 (SD = 1.98), a median of 8 and a range of 4-10. Fluency measured as spoken words/second in the Cinderella story, had a mean of 1.02 (SD = 0.50), median of 0.95 and a range of 0.24-2.30. The two measures were positively and reliably correlated (Spearman's rho = 0.30, p = 0.002), although the relatively small value of the correlation suggests that the two measures were sensitive to different aspects of fluency, hence our decision to include them both in the model.

Table 1 - Effect of non-cognitive factors on WAB-R fluency scores

Model term	**Coefficients**	**SE**	**t**	***p* value**
Intercept	9.04	1.68	5.39	<0.001
Gender	-0.48	0.40	-1.21	0.23
Age	-0.02	0.02	-1.12	0.26
Education	-0.02	0.07	-0.23	0.82
Years post-stroke	0.03	0.07	0.36	0.72
Years in speech-language therapy	-0.16	0.08	-2.04	0.04
Examiner's years of experience	0.02	0.02	0.80	0.43

-- Figure 2 about here --

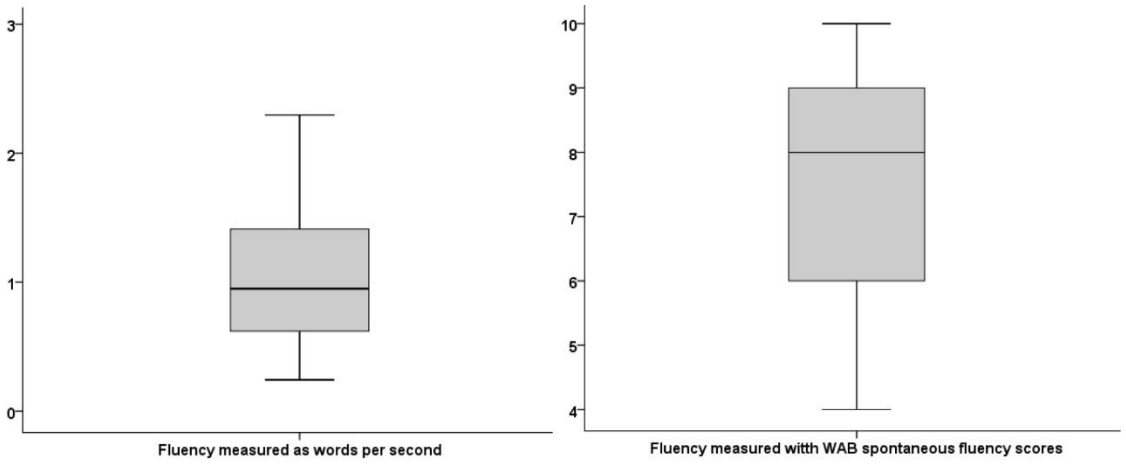

Figure 2- Box plots for the two measures of fluency in the participants (N = 112). The left panel shows fluency measured as words/second. Right panel shows fluency measured as WAB-R spontaneous fluency scores.

3.1.2. Measures of independent variables

Our choice of the latent independent variables was guided by the literature review discussed earlier. Table 2 lists all the latent and manifest independent variables in the model. The Object Naming subtest of WAB-R (Kertesz, 2006) elicits 20 common object names, with a maximum score of 3 for each correct response. Responses with phonemic paraphasias and following phonemic/tactile cues are scored as 2 and 1 respectively. Boston Naming Test (BNT)- short form (Goodglass et al., 2001) includes 20 pictures of both high and low frequency nouns, with varying phonological complexity. We re-scored the BNT responses of the 112 individuals using a more fine-grained measure: on each trial, if the picture was named correctly on first attempt and within 5 seconds of its presentation, the full score (3) was assigned. If the naming was delayed by more than 5 seconds, or if the correct name followed self-correction of an earlier erroneous response, the score 2 was assigned to the trial. If the individual only produced the correct name after the examiner provided semantic or phonological cues, the trial received a score of 1, and finally a score of 0 was assigned if the individual never arrived at the correct response. The third word production measure was the Verb Naming Test (VNT, Northwestern Assessment of Verbs and Sentences; Cho-Reyes & Thompson, 2012). The VNT elicits names of 22 action pictures, and includes intransitive, transitive and ditransitive verbs.

Syntactic production measures were obtained from the Cinderella story retell using CLAN analysis tools of TalkBank (MacWhinney, 2000). Five morphosyntactic measures were derived. The proportion of grammatical utterances was calculated as the number of grammatically accurate utterances divided by the total number of utterances. The Developmental Sentence Score (DSS, Lee & Canter 1971) provides an overall view of an individual's morphosyntactic abilities by taking into account

the use of eight syntactic constructions: indefinite pronouns or modifiers, personal pronouns, main verbs, embedded verbs, negatives, conjunctions, interrogative inversions, and the *wh*- question form. In addition to taking complexity and grammatical accuracy into account, DSS captures the use of function words and weighs grammatical forms based on the developmental sequence in which these are acquired. DSS is calculated over 50 utterances which must be complete (noun + verb), unique, consecutive, and not echoes of the conversation partner, in speakers whose production comprises at least 50% sentences (i.e., noun + verb). While DSS and proportion of grammatical utterances provide global measures of morphosyntactic ability, three specific morphosyntactic elements were also extracted, all of which are calculated as a percentage of the total number of words and use automated part of speech tagging of CLAN (MOR, MacWhinney, 2000). The percentage of verbs includes a count of all words tagged as a verb, including main verbs, participles, copulas and modals. Past tense is counted as verbs that have past tense affix, and prepositions include all words tagged as prepositions.

Comprehension was indexed by two scores in WAB-R: yes/no questions and word recognition. Yes/No questions involve 20 questions of varying complexity (e.g. Is your Name Smith? Do you peel a banana before you eat it?) with a score of 3 points for each correct answer. Verbal, gestural or eye-blink responses are allowed. Auditory word recognition involves pointing to common objects in the environment (e.g. window, ceiling), one's body parts, and pictured objects, colors, letters, and numbers , with a total of 60 items and 1 point for each correct identification. Objects, colors, letters and numbers are presented in a field of 6 items.

Working memory was represented by the span tasks in the AphasiaBank. These tests present participants with a list of words (repetition 1A), or word triplets (repetition 1B), or sentences (2A) that grow in length. For example, in repetition 1A, participant is first asked to repeat a list containing a single word, then two words, and so on up to eight words. Repetition 2B controls for reconstruction of materials in memory from the gist without remembering the exact utterance by including semantically

unlikely sentences such as “the bird was caught by the worm”. While span tests are efficient measures of working memory capacity in neurotypical individuals, in individuals with aphasia they are confounded by comprehension and repetition abilities. To account for the contribution of comprehension to working memory measured with these tests, we can simply define a dependency between the two (Figure 1). As for repetition, lexical repetition abilities in individual with aphasia are at least partially captured by picture naming abilities (Nozari & Dell, 2013; Nozari, Kittredge, Dell & Schwartz, 2010), which is represented by the latent variable “word production” in our model. So we also define a dependency between word production and working memory (Figure 1). What remains unaccounted for by comprehension and word production abilities can be attributed to working memory represented by this latent variable in the model. Finally, we included the latent variable “restarts” to capture the variability among participants in their tendencies to revise their speech. Two variables (repetition and retracing, Table 2) were used to represent the individuals’ tendency for repeating words and phrases which, if excessive, could create disfluent production.

In summary, the model (Figure 1) includes five independent latent variables, each represented by a set of manifest variables in reflective mode, and one independent latent variable (fluency) represented by two manifest variables. In addition to the links between the five independent variables and the dependent variable that directly tests its hypothesized influence on fluency, there are a number of interdependencies between the independent latent variables (comprehension predicting word production and working memory, word production predicting working memory, and working memory predicting syntactic abilities) to most accurately capture the cognitive architecture underlying speech fluency.

Table 2 - List of the independent latent variables, their corresponding manifest variables and measures, along with the group performance on each measure in our sample.

Latent variable	Manifest variables	Measurement	Mean	SD	Min	Max
a)Word production (in isolation)	a1. Object naming	WAB object naming , N = 20 (3 points for each correct response)	49.83	12.22	8	60
	a2. Picture naming of objects	BNT-short form. (see text for revised scoring criteria, 3 points for each correct response)	27.6	10.58	1	45
	a3. Picture naming of actions	VNT, Northwestern Assessment of Verbs. N = 22 (1 point for each correct response).	17.59	5.01	0	22
b) Syntactic production (in Cinderella story retell)	b1. Morphosyntactic accuracy	Percentage of utterances in the Cinderella story without a grammatical error	67	21	26	100
	b2. Morphosyntactic complexity	Developmental Sentence Score (see text for description)	12.21	3.34	3	20.61
	b3. % verbs	Proportion of verbs out of total words	17.91	4	4.72	28.85
	b4. % past tense	Proportion of regular past tense suffixes (-ed) out of verbs	7.11	4.75	0	15.93
	b5. % preposition	Proportion of prepositions out of total words	5.28	2.74	0	12.59
c) Working memory	c1. Word list recall	AphasiaBank - Repetition 1A (1-8 words)	3.41	2.04	1	8
	c2. List of lists recall	AphasiaBank - Repetition 1B (1-7 list of 3-word lists)	2.48	1.24	0	6
	c3. Sentence recall	AphasiaBank - Repetition 2A recall (1-6 sentences increasing in length)	39.61	22.39	0	84
	c4. Sentence recall (various sentence types)	AphasiaBank - Repetition 2B (containing semantically anomalous sentences or sentences that could create interference with reversible thematic roles)	58.12	28.43	0	88
d) Comprehension	d1. Command following	WAB yes/no commands	56.96	3.6	42	60
	d2. Word comprehension	WAB auditory word recognition	56.7	4.8	36	60
e) Restarts (in Cinderella story retell)	e1. Repetition	Instances of repetition of all or part of a spoken word without change	21.25	29.77	0	215
	e2. Retracing	Instances of phrase repetition with possible syntactic or lexical change; main idea is maintained	16.63	17.13	1	107

3.2. Model evaluation

One advantage of path modeling is that one can interactively test and revise the model to have the most accurate representation of the structure under investigation. The role of the outer model is

critical here. We often assume that our latent variables are coherent constructs best represented by the manifest variables we have selected for them. Inspection of the outer model allows us to empirically verify these assumptions and to revise the model accordingly. In other words, although most of the theoretical questions are answered by examining the inner model, a thorough examination of the outer model is necessary to ensure that a theoretically-plausible model with correct assumptions is being put to test. In this section, we discuss four critical steps involved in evaluating the appropriateness of the model. The first three involve assessing the outer model. The last step assesses the general fit of the model.

Step 1- Checking the unidimensionality of the manifest variables. Recall that a reflective latent variable is assumed to be the cause of its manifest variables. As such, a change in the latent variable is expected to affect all its manifest variables in the same direction, e.g., reduced working memory capacity should lead to worse performance on recalling words, word lists and sentences. This is called *unidimensionality* and represents a measure of composite reliability for the latent variable. If performance on one of these tasks actually improves with lower working memory scores, the effect of the latent variable on its manifest variables is no longer unidimensional. In that case, the model needs to be revised to determine the reason behind the violation of unidimensionality. There are several indices for quantifying unidimensionality, including Cronbach's alpha and the first eigenvalue of the indicators' correlation matrix (see Segars, 1997 for more discussions). We report Dillon-Goldstein's rho (DG rho), which focuses on the variance of the sum of variables that contribute to a given latent variable. As a rule of thumb, a DG rho of 0.7 or higher is taken to represent good unidimensionality for a latent variable. DG rho's for the latent variables in our fluency model were as follows: fluency = 0.79, comprehension: 0.83, word production: 0.92, syntactic production: 0.82, working memory: 0.93, and restarts: 0.93. Since all DG rho's are high, it seems that representing our latent variables by their manifest variables is appropriate in the model.

Step 2 - Checking the loadings of the manifest variables. The loadings reflect the correlations between a latent variable and its manifest variables. Each manifest variable is considered to reflect the contribution of the latent variable plus noise or task-specific factors. The relative contribution of the latent variable vs. noise/task-specific factors is reflected in the loading for that manifest variable. Ideally, one hopes that more variance in the manifest variable is explained by the latent variable than by noise/task-specific factors. This can be indexed by a loading of 0.7, which implies that about 50% of the variability in the manifest variable is explained by the latent variable. Because cognitive measures especially in neuropsychological testing are often noisy, we opt for accepting manifest variables with loading of 0.6 and greater. Indicators with considerably lower loadings may have to be reconsidered.

The diagonal of Figure 3 contains cells that show the loading of each set of manifest variables on the latent variable they are meant to represent. Each bar in a cell is a separate manifest variable. For example, the first cell contains two bars representing scores from the two comprehension tests and their loadings on the comprehension latent variable, and so on. All the loadings are higher than 0.6, except for the words/sec measure of fluency which has a loading of 0.48 (see Table B1 in the Appendix B for all loading and crossloading values). This value is borderline, and one approach is to drop this variable from the model. On the other hand, we expected WAB-R fluency scores and word/sec to be sensitive to different aspects of fluency, and this was our reason for including both variables in the model. Therefore, we retain both variables, keeping in mind that the WAB-R fluency scores influence the results more heavily in this model. We then run a second model with only words/second as the fluency measure and compare the results.

-- Figure 3 about here –

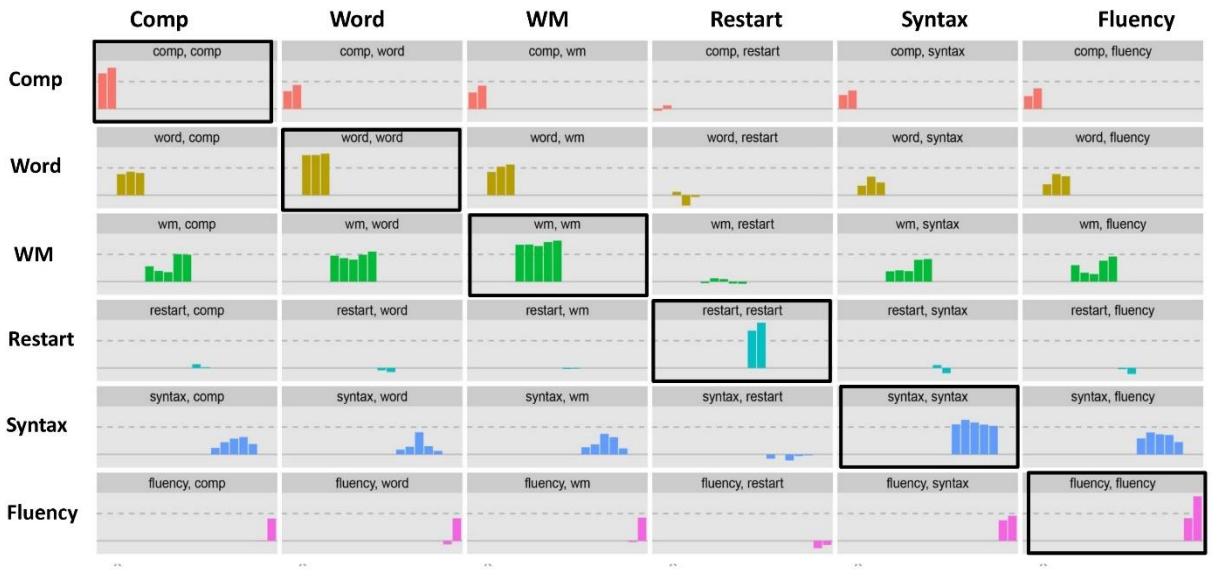

Figure 3 - Loading (diagonal cells) and cross-loading (off-diagonal cells) of the manifest variables under each latent variable on that variable and other latent variables in the model. The dashed line marks the criterion 0.6. The diagonal (black-outlined boxes) show that the tests that are meant to measure a certain variable indeed have the highest loading on that variable. See Table B1 in Appendix B for the values of loading and crossloading for each test.

Step 3 - Checking the cross loadings of the manifest variables. Ideally, each manifest variable should reflect significant contribution of one latent variable, and as little contribution of the other latent variables as possible. This can be tricky for tasks reflecting cognitive functions, as many such tasks tap into several cognitive abilities. It is nevertheless important to ensure that the loading of a manifest variable is indeed highest on the latent variable it is assumed to represent. The off-diagonal cells in Figure 3 show the crossloadings of the manifest variables on other latent variables, and Table B1 in Appendix B lists the values of all the crossloadings. Using these values, we can check if a variable such as VNT which involves producing a verb, is correctly placed under word production, or if it is better suited

as representing syntactic production abilities. Table B1 shows that VNT loads 0.88 on word production and 0.41 on syntactic production. We thus conclude that its placement under word production is the more appropriate choice, in keeping with the theoretical arguments we discussed in section 1.3. This approach can be used to revise the model and determine the best arrangement of variables.

Step 4 - Checking the general fit of the model. The first three steps evaluated the quality of the model by examining the outer model. These checks are used to revise the model, by potentially dropping manifest variables or reconsidering the latent variable constructs. The revised models must be checked through the same steps again. Once a viable model is obtained, the fourth check is performed to assess the overall fit of the model, which also includes the inner model. Two indices are used for this purpose: (a) R^2 is reported for the latent variable of interest and, similar to simple regression models, indicates the amount of variance explained by the independent latent variables, i.e., predictors. (b) The model also generates an index of the goodness of fit (GoF). The problem is that, similar to multi-level models, there is no absolute criterion to measure the overall quality of the model. For this reason, we will focus on R^2 instead. The R^2 in the current model was 0.44, meaning that 44% of variance in fluency was explained by the model.

3.3. Interpreting the model's results

Figure 4 shows the parameter values estimated by the pls-pm for both the inner and the outer models. While most of the model evaluation work focused on the outer model, it is the inner model that answers questions of theoretical interest to us. The output of the inner model is similar to that of any generalized linear model with parameter estimates (path coefficients) and their corresponding t statistics. Since pls-pm does not rely on any distributional assumptions, significance levels of path

coefficients are best estimated through resampling methods. We report 95% confidence intervals obtained via bootstrapping[1] with 10,000 iterations.

Table 3 presents the results of the bootstrapping procedure. Two sets of effects are reported. The upper part of the table reports the *direct effects*. This is similar to a simple regression, where the effect of an independent variable is reported on a dependent variable that it is immediately connected to. The table shows that our assumptions about the interdependence of the independent variables were correct: the 95% CI for the effect of comprehension over word production, the effects of both comprehension and word production on the working memory scores, and the effect of working memory on syntactic production do not contain zero, suggesting that all these effects are significant at $\alpha = 0.05$. Of the independent variables included in the model, the only one that has a statistically-reliable direct effect on fluency is syntactic production.

-- Figure 4 about here --

[1] Bootstrapping is a technique for estimating the sampling distribution of a statistic using random sampling with replacement.

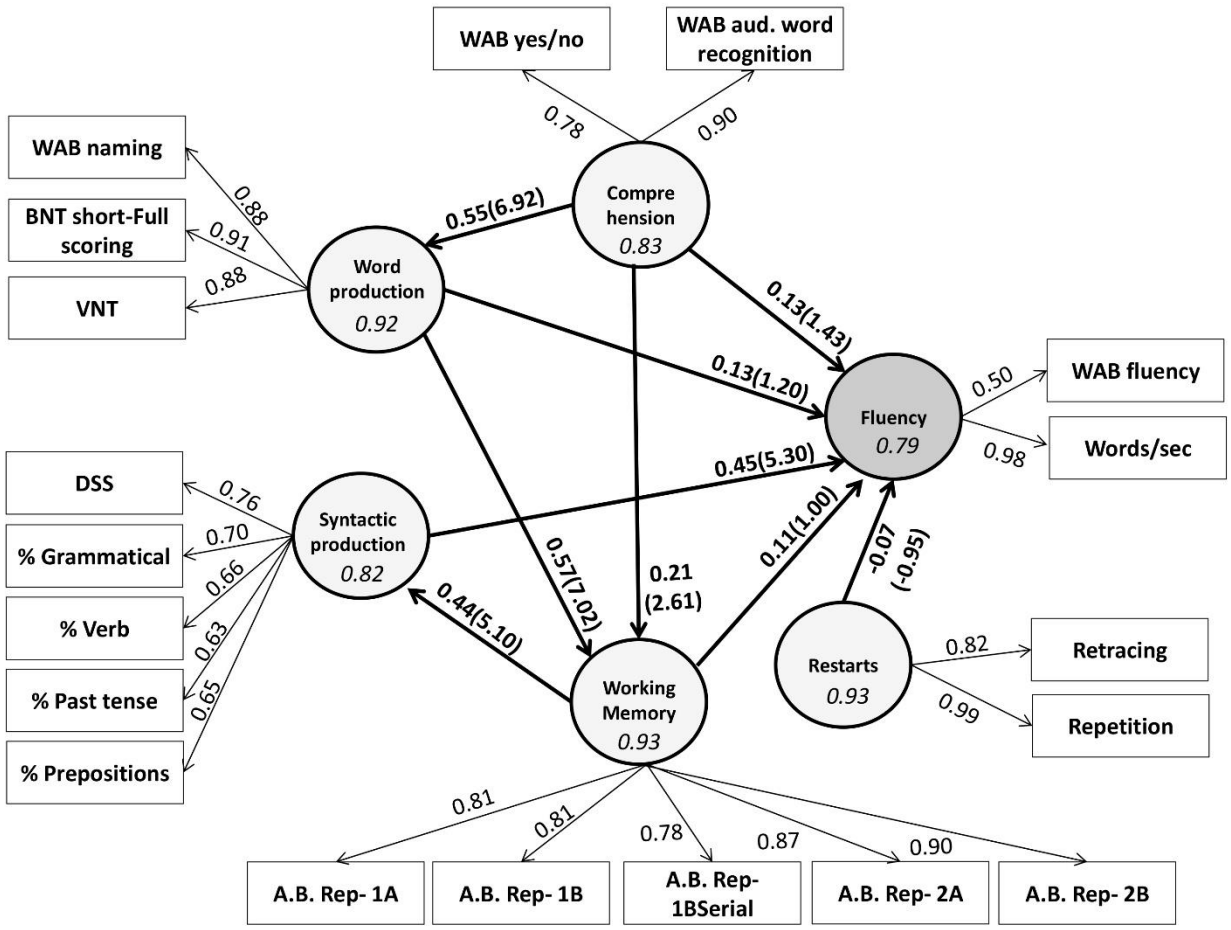

Figure 4- Parameter estimates for the inner and outer models in the fluency pls-pm. The values on the arrows from the latent variables (circles) to manifest variables (rectangles) are factor loadings. The values on the arrows between the latent variables are path coefficients. The values inside the parentheses next to the path coefficients are their corresponding t statistics. See Table 3 for 95%CIs obtained through bootstrapping. The values inside the circles are DG rho's.

Table 3- Results of the inner model, obtained by 10,000 bootstrapping iterations.

Effects	Original path coefficient	Mean bootsrapped coefficient	SE	95% CI - Lower bound	95% CI - Upper bound
Direct effects					
comp -> word	0.55	0.55	0.07	0.42	0.68
comp -> wm	0.21	0.22	0.08	0.07	0.38
word -> wm	0.57	0.57	0.08	0.41	0.71
wm -> syntax	0.44	0.45	0.06	0.33	0.57
comp -> fluency	0.13	0.13	0.10	-0.09	0.31
word -> fluency	0.13	0.13	0.12	-0.12	0.35
wm -> fluency	0.11	0.11	0.13	-0.15	0.35
restart -> fluency	-0.07	-0.07	0.08	-0.23	0.10
syntax -> fluency	0.45	0.43	0.12	0.21	0.60
Total effects (direct +indirect effects)					
comp -> word	0.55	0.55	0.07	0.42	0.68
comp -> wm	0.53	0.54	0.06	0.41	0.65
comp -> syntax	0.23	0.24	0.05	0.16	0.34
word -> wm	0.57	0.57	0.08	0.41	0.71
word -> syntax	0.25	0.26	0.05	0.17	0.36
wm -> syntax	0.44	0.45	0.06	0.33	0.57
comp -> fluency	0.36	0.36	0.12	0.20	0.51
word -> fluency	0.30	0.30	0.11	0.13	0.46
wm -> fluency	0.30	0.30	0.14	0.03	0.55
restart -> fluency	-0.07	-0.07	0.08	-0.23	0.10
syntax -> fluency	0.45	0.43	0.12	0.21	0.60

Note that we now have a dilemma on our hands. A variable such as working memory has a significant effect on syntactic production, which in turn has a significant effect on fluency, but the direct effect of working memory on fluency is not statistically significant. Should we conclude that working memory affects fluency or not? The solution to this problem in the pls-pm framework is provided by the inspection of *indirect effects*, which are presented in the lower part of Table 3. An indirect effect characterizes the influence of one variable over another through intermediate variables. The indirect effects are calculated as the product of the path coefficients involved in the indirect path. For example, the indirect effect of working memory on fluency (Figure 1) is calculated as the path coefficient of wm

→ syntax (0.44) times the path coefficient of syntax → fluency (0.45), returning an indirect effect of 0.19 for wm → fluency. Once we have the values of the indirect effects, we can calculate the *total effect* of an independent variable on the dependent variable of interest, by summing its direct and indirect effects on that variable. In this case, the total effect of wm → fluency would be calculated as 0.11 (taken from the upper part of Table 3) plus 0.19, returning a total effect of 0.30 (the effect reported in the lower part of Table 3). Figure 5 shows the direct and indirect effects for each of the five independent latent variables on our dependent variable of interest, fluency. Looking at the total effects in Table 3, we now see that the 95% CIs for the effects of four out of five independent variables on fluency do not contain zero, suggesting a reliable influence of comprehension, single-word production, and working memory, in addition to syntactic production abilities, on fluency.

-- Figure 5 about here --

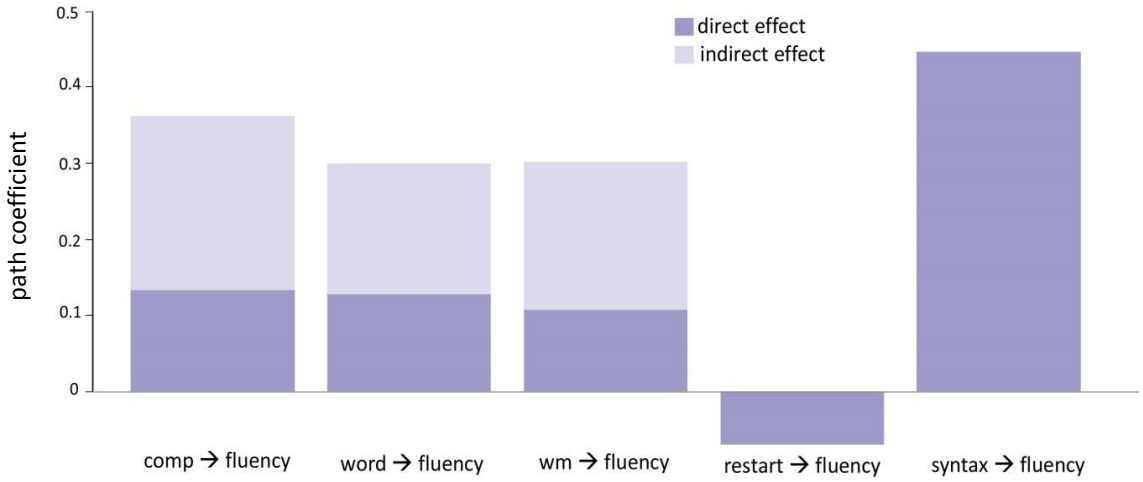

Figure 5 - Direct and indirect effects of the five independent variables on the dependent variable fluency. The full height of the bars shows the total effect.

To summarize, the results of the pls-pm in a model with fluency represented by both WAB-R spontaneous fluency scores and words/second revealed reliable effects of comprehension, working memory, single-word production abilities, and syntactic production abilities on fluency. Of these, however, only syntactic production abilities had a reliable direct effect on fluency (see Figure 5). Recall that during model evaluation, the loading of WAB-R fluency scores on the fluency latent variable was much higher than that of words/second. Recall also that WAB-R spontaneous fluency scores inherently contain an evaluation of syntactic abilities. Thus, it is fair to question whether the obtained results are biased towards finding a stronger influence of syntactic production abilities (compared to any other variable) on fluency. To address this concern, we ran the model again, this time including only words/second as the manifest variable representing fluency. This variable, unlike WAB-R fluency scores, does not favor good syntactic abilities. In other words, "He want go my store" and "He goes to the store" would receive the same score (5 words each) if spoken within the same time window. Thus, if the reliable effect of syntactic abilities on fluency obtained in the previous model was due to the use of WAB-R fluency scores as the dependent variable, we could expect a different pattern when the dependent variable is changed to words/second.

The same checks were applied to the reduced model. The model passed all the criteria and its R^2 was 0.33, showing a lower explanatory power than the previous model ($R^2 = 0.44$). Critically though, the 95% CI for the effect of syntax → fluency was 0.47 to 0.74, which did not contain zero, confirming the statistically-reliable effect of syntactic production on predicting fluency in the absence of a scoring system that is biased towards giving more weight to syntactic abilities. In fact, in the reduced model with only words/sec as the outcome variable, syntactic production was the only independent variable whose total effect was a reliable predictor of fluency (see Table B2 in the Appendix B for the full results of this analysis).

4. Discussion

This paper followed two goals: (a) to present an overview of path modeling, a statistical approach to testing complex relationships between multiple variables each represented by several observed measures, and to show its usefulness in exploring cognitive and neuropsychological data; and (b) to employ the model to examine the origin of nonfluency in post-stroke aphasia. In fulfilling the first goal, we discussed the advantages of pls-pm over simple regression methods, followed by a step-by-step walk-through of how to set up, evaluate, and interpret a path model. The general framework discussed here is easily adaptable to various kinds of neuropsychological or cognitive phenomena with multi-factorial origin. For example, one could study the contribution of inhibitory control and working memory to production of agreement attraction errors in neurotypical children and adults by constructing a model with agreement errors as its dependent variable, and inhibitory control and working memory as independent latent variables each represented by several manifest variables. Similarly, the origin of language impairment in non-aphasic syndromes, such as semantic dementia, or natural conditions such as aging, can be probed by constructing a model that includes latent variables of possible interest such as comprehension, working memory, repetition, etc. Importantly, path modeling presents an alternative to routine statistical approaches that often overlook the interdependent network of factors that underlie a variable of interest.

In addressing the second goal, we reviewed several bodies of literature that have proposed contradictory evidence for the association between disfluent/nonfluent speech and various cognitive abilities. Importantly, none of these studies have simultaneously accounted for all the variables included in the current study.

The results unequivocally supported an effect of syntactic production abilities on nonfluency. This finding indicates a robust connection between agrammatism and nonfluency in aphasia (also see

Bonner & Grossman, 2010 for the same link in primary progressive aphasia; but see Saffran et al., 1989; Thompson et al., 2012). Critically, word production, comprehension and working memory also had a significant, albeit indirect, effect on fluency. There is an important message in this finding: impairment in *any* of these abilities could cause disfluency. For example, impairment in working memory alone could lead to nonfluency, even though this impairment will also manifest as a breakdown in morphosyntactic abilities. In other words, working memory provides a common cause for both morphosyntactic problems and nonfluency. This, in turn, has implications for language rehabilitation. While syntactic impairment and nonfluency are correlated, the focus of treatment should be on a variable that mediates both.

One remaining variable (restarts) did not show a reliable effect on fluency. Recall that this variable was used as an index of monitoring. However, as acknowledged in the earlier sections, revisions and retraces are sub-optimal indices of monitoring abilities. For one thing, the main monitoring ability (i.e., detection of errors) is not well captured by these measures, as both repetition and retracing may happen on already correct words. Moreover, there are reasons other than monitoring that may cause speakers to repeat a word or a phrase. For example, problems with lexical and phonological retrieval may lead to repetitions (Fraundorf & Watson, 2014). Thus better measures of monitoring are required before a role of impaired monitoring in nonfluent production in PWA can be refuted.

While the study succeeded in its goal to take the first step in elucidating the cognitive architecture of nonfluency in aphasia, there are several caveats that must be improved upon in future studies. The R^2 for the model with both fluency measures was 0.44, suggesting that only about half the variance in the data is accounted for by the variables included in this model. Model performance can be improved in two ways: (a) selection of better tests for the existing variables, and (b) inclusion of more variables. As alluded to in the earlier sections, AphasiaBank is an excellent resource for large, consistently administered and coded data in PWA. However, the tests are not always ideal for

representing the cognitive processes targeted in psycholinguistic and cognitive theories. For example, ideally, we would have liked to represent semantic comprehension with tests that involved no language. Similarly, working memory could have been represented using measures that distinguish between the dissociable semantic and phonological working memory (Martin, Shelton, & Yaffee, 1994). Future studies should aim at exploring the contribution of each variable indicated in this model using manifest variables that more closely capture the cognitive mechanisms represented by that variable. In a similar vein, some of the independent latent variables can be dissected into finer sections informed by relevant theories. For example, while we included syntactic production as one latent variable, various syntactic operations can perhaps be dissociated and may have different bearings on nonfluency. Due to the nature of our data, the current variable weighs morphosyntactic production more heavily than other aspects of syntactic operations, such as differences in the number of arguments associated with a verb and its effect on the difficulty of verb retrieval (e.g., Kim and Thompson 2004). Fortunately, the pls-pm framework can be easily applied to further break down each latent variable in the model. For example, with a dataset rich in various syntactic measures, one could define more fine-grained latent variables representing different aspects of syntactic processing, each with their own set of manifest variables, and test which aspect contributes most significantly to nonfluency.

Finally, there are also certain variables that are very likely to affect fluency, but were not included in the current model because AphasiaBank either contained no measure for such variables or the scores were missing for a large number of patients on the existing measures. Two such variables are particularly noteworthy: apraxia of speech, and executive control abilities.

Apraxia of speech. Apraxia of speech sometimes follows lesion to the left hemisphere, and is characterized by phonological and phonetic errors, distorted prosody, visible struggle in controlling articulators (also called groping) and inconsistency in errors. The origin of speech apraxia is

controversial, but most likely lies somewhere between phonological encoding and motor implementation (e.g., Hickok, 2012; McNeil, Robin, & Schmidt, 2009; Laganaro, 2012), and possibly concerns a specific stage of articulatory planning/programming (Galluzzi, Bureca, Guariglia, & Romani, 2015). Apraxia of speech typically causes nonfluency, which has been attributed by some to the loss of stored motor programs (e.g., Varley & Whiteside, 2001) and by others to a procedural impairment in online computation of articulatory plans (e.g., Ziegler, 2011). As such, it is an important variable to include in nonfluency studies, the scope of which reaches past phonological encoding.

Executive control abilities. Executive control abilities are a family of functions that regulate other cognitive operations, such as language production (Nozari & Novick, in press). A large number of studies show interference in producing a target word among semantically-related (e.g., Costa, Alario, & Caramazza, 2005; Schnur et al., 2006; Schriefers, Meyer, & Levelt, 1990, but see Mahon et al., 2007) and phonologically-related items (Breining, Nozari, & Rapp, 2016; Nozari et al., 2016). There is both behavioral and neural evidence that resolving such competition requires executive control (Shao, Meyer, & Roelofs, 2013; Shao, Roelofs, & Meyer, 2012; Schnur et al., 2009). At the sentence level too, accuracy and timing of processing have been shown to depend on executive control abilities and their neural correlates, especially the left lateral prefrontal cortex, in both sentence comprehension (Nozari, Mirman, & Thompson-Schill, 2016; Nozari, Trueswell, & Thompson-Schill, 2016) and sentence production (Arnold & Nozari, 2017; Nozari, Arnold, & Thompson-Schill, 2014; see Nozari & Thompson-Schill, 2015 for a review).

These and similar studies suggest that impaired executive abilities could render the processes involved in word and sentence production less efficient, by prolonging the time required to resolve

competition between alternatives, and in some cases failing to resolve such competition altogether, as has been implied in certain individuals with dynamic aphasia (Robinson, Blair, & Cipolotti, 1998; Robinson, Shallice, Bozzalli & Cipolotti, 2010), as well as semantic aphasia (Jefferies, Patterson, & Lambon Ralph, 2008). To directly test the relationship between executive control abilities and fluency, Nozari and Schwartz (2012) tested three PWA, two nonfluent and one fluent, matched on comprehension and single-word production abilities, and showed that the two nonfluent individuals were significantly impaired in their ability to efficiently switch between tasks (a hallmark of executive control), while the fluent individual showed no impairment in this task when compared to neurotypical age-matched control. Thus indices of executive control, including inhibitory control and task-switching, seem quite relevant to fluency. Unfortunately, AphasiaBank contains no such measures, but future research must focus on integrating these measures in models of speech fluency.

Finally, the study of nonfluency in aphasia can benefit from more comprehensive measures. A particularly appealing candidate is discourse measures, such as coherence and cohesion. Indeed, in a recent study Arnold and Nozari (2017) showed that prefrontally-mediated operations, especially working memory, were critical to maintaining discourse connectivity. Thus, it is reasonable to assume that impairment of such abilities could lead to disruption in discourse, which could in turn contribute to nonfluency.

5. Conclusion

This work took the first step in illuminating the multi-faceted origin of nonfluency in PWA using path modeling, and identified word production, comprehension, and working memory deficits as contributing to nonfluency, in addition to syntactic impairment which had a strong and direct impact on speech fluency. We urge future studies to focus on speech apraxia, monitoring and executive control abilities as the three remaining variables with the greatest likelihood of contributing to nonfluent

aphasia. More generally, we demonstrated that a path model is an excellent tool for exploring complex neuropsychological symptoms of multi-factorial origin.

6. Acknowledgments

This work was supported by the NSF grant 1631993 to Nazbanou Nozari, and by the NIH grant R01-DC008524 to Brian MacWhinney for the AphasiaBank. We thank all the researchers and individuals with aphasia who support the AphasiaBank by publicly sharing their data, and Brain MacWhinney and his group for supporting and maintaining this valuable database.

6. References

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological bulletin*, *135*(1), 50. doi: 10.1037/a0014411

Andreetta, S., & Marini, A. (2015). The effect of lexical deficits on narrative disturbances in fluent aphasia. *Aphasiology, 29*(6), 705-723. doi: 10.1080/02687038.2014.979394

Arnold, J. E., & Nozari, N. (2017). The effects of utterance timing and stimulation of left prefrontal cortex on the production of referential expressions. *Cognition*, *160*, 127-144. doi: 10.1016/j.cognition.2016.12.008

Bastiaanse, R., & Jonkers, R. (1998). Verb retrieval in action naming and spontaneous speech in agrammatic and anomic aphasia. *Aphasiology*, *12*(11), 951-969. doi:10.1080/02687039808249463

Benson, D. F. (1967). Fluency in Aphasia: Correlation with Radioactive Scan Localization. *Cortex, 3*(4), 373-394. doi: http://dx.doi:10.1016/S0010-9452(67)80025-X

Bird, H., & Franklin, S. (1996). Cinderella revisited: A comparison of fluent and non-fluent aphasic speech. *Journal of Neurolinguistics*, *9*(3), 187-206. doi: 10.1016/0911-6044(96)00006-1

Bock, K. (1995). Sentence Production: From Mind to Mouth. In J. L. Miller & P. D. Eimas (Eds.), *Speech, Language and Communication* (pp. 181-216). San Diego: Academic Press.

Bonner, M. F., Ash, S., & Grossman, M. (2010). The new classification of primary progressive aphasia into semantic, logopenic, or nonfluent/agrammatic variants. *Current neurology and neuroscience reports*, *10*(6), 484-490. 10.1007/s11910-010-0140-4

Breining, B., Nozari, N., & Rapp, B. (2016). Does segmental overlap help or hurt? Evidence from blocked cyclic naming in spoken and written production. *Psychonomic bulletin & review*, *23*(2), 500-506. doi: 10.3758/s13423-015-0900-x

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81–116). Hove, UK: Psychology Press.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and brain Sciences*, *22*(01), 77-94. doi:10.1017/s0140525x99001788

Caspari, I., Parkinson, S. R., LaPointe, L. L., & Katz, R. C. (1998). Working memory and aphasia. *Brain and cognition*, *37*(2), 205-223. doi: 10.1006/brcg.1997.0970

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology, 26*(10), 1250-1277. doi:10.1080/02687038.2012.693584

Costa, A., Alario, F. X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review*, *12*(1), 125-131. doi: 10.3758/bf03196357

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of memory and language*, *27*(2), 124-142.

Dell, G. S., Nozari, N., & Oppenheim, G. M. (2014). Lexical access: Behavioral and computational considerations. In V. Ferreira, M. Goldrick, & M. Miozzo (Eds.), *The Oxford Handbook of Language Production* (pp. 88-104). Oxford: Oxford University Press.

Edwards, S. (1995). Profiling fluent aphasic spontaneous speech: A comparison of two methodologies. *International Journal of Language & Communication Disorders*, *30*(3), 333-345. doi: 10.3109/13682829509021446

Edwards, S. (1998). Single words are not enough: Verbs, grammar and fluent aphasia. *International journal of language & communication disorders*, *33*(S1), 190-195. doi: /10.3109/13682829809179421

Edwards, S., & Bastiaanse, R. (1998). Diversity in the lexical and syntactic abilities of fluent aphasic speakers. *Aphasiology*, *12*(2), 99-117. doi: 10.1080/02687039808250466

Edwards, S., Bastiaanse, R., & Kiss, K. (1994). Fluent aphasia and grammatical complexity in three languages. *Brain and Language, 47*, 414-416.

Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning & Verbal Behavior, 19*, 624–634. doi: 10.1016/s0022-5371(80)90672-6

Faroqi-Shah, Y., & Thompson, C. K. (2003). Effect of lexical cues on the production of active and passive sentences in Broca's and Wernicke's aphasia. *Brain and Language, 85*(3), 409. doi.org/10.1016/s0093-934x(02)00586-2

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*, 92-107. doi: 10.2307/1937887

Feyereisen, P., Pillon, A., & de Partz, M. (1991). On the measures of fluency in the assessment of spontaneous speech production by aphasic subjects. *Aphasiology, 5*(1), 1-21. doi: 10.1080/02687039108248516

Fillmore, C. (1979). *On fluency*. New York: Academic Press.

Fraundorf, S. H., & Watson, D. G. (2014). Alice's adventures in um-derland: Psycholinguistic sources of variation in disfluency production. *Language, Cognition and Neuroscience*, *29*(9), 1083-1096. doi: 10.1080/01690965.2013.832785

Galluzzi, C., Bureca, I., Guariglia, C., & Romani, C. (2015). Phonological simplifications, apraxia of speech and the interaction between phonological and phonetic processing. *Neuropsychologia*, *71*, 64-83.doi: 10.1016/j.neuropsychologia.2015.03.007

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 9, pp. 133-177). New York: Academic Press.

Goodglass, H., & Kaplan, E. (1983). Major aphasic syndromes and illustrations of test patterns *The assessment of aphasia and related disorders* (pp. 74-100). Philadelphia: Lea and Febiger.

Goodglass, H., Christiansen, J. A., & Gallagher, R. (1993). Comparison of morphology and syntax in free narrative and structured tests: fluent vs. nonfluent aphasics. *Cortex, 29*(3), 377-407. doi: 10.1016/s0010-9452(13)80250-x

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *Boston Diagnostic Aphasia Examination, 3rd Edition*. Phidelphia: Lippincott Williams and Wilkins.

Gordon, J. K. (1998). The fluency dimension in aphasia. *Aphasiology, 12*(7/8), 673-688.

Gordon, J. K. (2006). A quantitative production analysis of picture description. *Aphasiology, 20*(2-4), 188-204. doi: 10.1080/02687030500472777

Griffin, Z. M., & Ferreira, V. S. (2006). Properties of spoken language production. In Traxler, M. & Gernsbacher, M. (Eds.) *Handbook of psycholinguistics* (pp. 21-59). Cambridge, MA: Academic Press.

Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes*, *21*(1-3), 181-204. doi:10.1080/01690960400002117

Hartsuiker, R. J., & Notebaert, L. (2009). Lexical access problems lead to disfluencies in speech. *Experimental psychology*. doi:10.1027/1618-3169/a000021

Hartsuiker, R., & Kolk, H. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology, 42*, 113-157. doi: 10.1006/cogp.2000.0744

Haynes, W. O., & Hood, S. B. (1978). Disfluency changes in children as a function of the systematic modification of linguistic complexity. *Journal of Communication Disorders*, *11*(1), 79-93. doi: 10.1016/0021-9924(78)90055-2

Hennessey, N. W., Nang, C. Y., & Beilby, J. M. (2008). Speeded verbal responding in adults who stutter: Are there deficits in linguistic encoding? *Journal of Fluency Disorders, 33*(3), 180-202. doi: http://dx.doi:10.1016/j.jfludis.2008.06.001

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13*(2), 135-145. doi: 10.1038/nrn3158

Holland, A. L., Fromm, D., & Swindell, C. S. (1986). The labeling problem in aphasia: an illustrative case. *Journal of Speech and Hearing Disorders, 51*(2), 176-180. doi:10.1044/jshd.5102.176

Howell, P., Au-Yeung, J., & Sackin, S. (1999). Exchange of stuttering from function words to content words with age. *Journal of Speech, Language, and Hearing Research*, *42*(2), 345-354. doi: 10.1044/jslhr.4202.345

Howes, D. (1964). Application of teh word-frequency concept to aphasia. In A. V. S. de Reuck & M. O'Connor (Eds.), *Disorders of Language* (pp. 47-78). London: J. & A. Churchill.

Howes, D., & Geschwind, N. (1964). The brain and disorders of communication: quantitative studies of aphasic language. *Research publications-Association for Research in Nervous and Mental Disease, 42*, 229-244.

Ivanova, M. V., Dragoy, O. V., Kuptsova, S. V., Ulicheva, A. S., & Laurinavichyute, A. K. (2015). The contribution of working memory to language comprehension: differential effect of aphasia type. *Aphasiology, 29*(6), 645-664. doi: 10.1080/02687038.2014.975182

Jacks, A., & Haley, K. L. (2015). Auditory Masking Effects on Speech Fluency in Apraxia of Speech and Aphasia: Comparison to Altered Auditory Feedback. *Journal of Speech Language and Hearing Research, 58*(6), 1670-1686. doi: 10.1044/2015_jslhr-s-14-0277

Jefferies, E., Patterson, K., & Ralph, M. A. L. (2008). Deficits of knowledge versus executive control in semantic cognition: Insights from cued naming. *Neuropsychologia, 46(*2), 649-658. doi: 10.1016/j.neuropsychologia.2007.09.007

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149. doi:10.1037//0033-295x.99.1.122

Kerschensteiner, M., Poeck, K., & Brunner, E. (1972). The fluency-non fluency dimension in the classification of aphasic speech. *Cortex, 8*(2), 233-247. doi:10.1016/s0010-9452(72)80021-2

Kertesz, A. (1982). *Western Aphasia Battery*. New York: Grune Stratton.

Kertesz, A. (2006). *Western Aphasia Battery-Revised*. San Antonio, TX: Pearson

Kim, M., & Thompson, C. K. (2000). Patterns of comprehension and production of nouns and verbs in agrammatism: Implications for lexical organization. *Brain and language*, *74*(1), 1-25. doi: 10.1006/brln.2000.2315

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. New York: Guilford publications.

Kok, P., van Doorn, A., & Kolk, H. (2007). Inflection and computational load in agrammatic speech. *Brain and Language, 102*(3), 273-283. doi: 10.1016/j.bandl.2007.03.001

Kolk, H. H. J., & Heeschen, C. (1992). Agrammatism, paragrammatism and the management of language. *Language and Cognitive Processes, 7*, 89-129. doi: 10.1080/01690969208409381

Kolk, H. H. J., & Heeschen, C. (1996). The malleability of agrammatic symptoms: A reply to Hesketh and Bishop. *Aphasiology, 10*, 81-96.

Kreindler, A., Mihailescu, L., & Fradis, A. (1980). Speech fluency in aphasics. *Brain Lang, 9*(2), 199-205.

Laganaro, M. (2012). Patterns of impairments in AOS and mechanisms of interaction between phonological and phonetic encoding. *Journal of Speech, Language, and Hearing Research*, *55*(5), S1535-S1543.doi: 10.1044/1092-4388(2012/11-0316)

Lee, L. L., & Canter, S. M. (1971). Developmental Sentence Scoring: A Clinical Procedure for Estimating Syntactic Development in Children's Spontaneous Speech. *Journal of Speech and Hearing Disorders, 36*(3), 315-340. doi: 10.1044/jshd.3603.315

Levelt, W. (1999). Producing spoken language: a blueprint of the speaker. In C. M. Brown & P. Hagoort (Eds.), *The Neurocognition of Language* (pp. 83-122). New York: Oxford University Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lohmöller, J. B. (2013). *Latent variable path modeling with partial least squares*. Berlin: Springer Science & Business Media.

Luzzatti, C., Raggi, R., Zonca, G., Pistarini, C., Contardi, A., & Pinna, G. D. (2002). Verb–noun double dissociation in aphasic lexical impairments: The role of word frequency and imageability. *Brain and language*, *81*(1), 432-444. doi:10.1006/brln.2001.2536

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates

Macwhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for Studying Discourse. *Aphasiology, 25*(11), 1286-1307. doi: 10.1080/02687038.2011.589893

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 503. doi: 10.1037/0278-7393.33.3.503

Marchina, S., Zhu, L. L., Norton, A., Zipse, L., Wan, C. Y., & Schlaug, G. (2011). Impairment of Speech Production Predicted by Lesion Load of the Left Arcuate Fasciculus. *Stroke, 42*(8), 2251-2256. doi: 10.1161/strokeaha.110.606103

Martin, R. C. (1987). Articulatory and phonological deficits in short-term memory and their relation to syntactic processing. *Brain and Language*, *32*(1), 159-192. doi: 10.1016/0093-934x(87)90122-2

Martin, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*, *33*(1), 83. doi: 10.1006/jmla.1994.1005

McCarthy, R., & Warrington, E. K. (1985). Category specificity in an agrammatic patient: The relative impairment of verb retrieval and comprehension. *Neuropsychologia*, *23*(6), 709-727. doi:10.1016/0028-3932(85)90079-x

McNeil, M.R., Robin, D.A., Schmidt, R.A. (2009). Apraxia of speech. In M.R. McNeil (Ed.), *Clinical Management of Sensorimotor Speech Disorders* (pp. 249–268). New York: Thieme.

Melnick, K. S., & Conture, E. G. (2000). Relationship of length and grammatical complexity to the systematic and nonsystematic speech errors and stuttering of children who stutter. *Journal of Fluency Disorders, 25*(1). doi: http://dx.doi:10.1016/S0094-730X(99)00028-5

Nozari, N., Arnold, J. E., & Thompson-Schill, S. L. (2014). The effects of anodal stimulation of the left prefrontal cortex on sentence production. *Brain stimulation*, *7*(6), 784-792. doi: 10.1016/j.brs.2014.07.035

Nozari, N., & Dell, G. S. (2013). How damaged brains repeat words: A computational approach. *Brain and language*, *126*(3), 327-337. doi:10.1016/j.bandl.2013.07.005

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive psychology*, *63*(1), 1-33. doi:10.1016/j.cogpsych.2011.05.001

Nozari, N., Freund, M., Breining, B., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and repair in word production. *Language, Cognition and Neuroscience*, *31*(7), 886-903. doi:10.1080/23273798.2016.1157194

Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of memory and language*, *63*(4), 541-559. doi:10.1016/j.jml.2010.08.001

Nozari, N., & Novick, J. (in press). Monitoring and control in language production. *Current Directions in Psychological Science.* doi: 10.1177/0963721417702419

Nozari, N., Mirman, D., & Thompson-Schill, S. L. (2016). The ventrolateral prefrontal cortex facilitates processing of sentential context to locate referents. *Brain and language*, *157*, 1-13. doi: 10.1016/j.bandl.2016.04.006

Nozari, N. & Thompson-Schill, S.L. (2015). Left Ventrolateral Prefrontal Cortex in Processing of Words and Sentences. In G. Hickok & S.L. Small (Eds.), *The Neurobiology of Language (pp. 569-588).* Waltham, MA: Academic Press.

Nozari, N., Trueswell, J. C., & Thompson-Schill, S. L. (2016). The interplay of local attraction, context and domain-general cognitive control in activation and suppression of semantic distractors during sentence comprehension. *Psychonomic bulletin & review*, *23*(6), 1942-1953. doi: 10.3758/s13423-016-1068-8

Nozari, N., & Schwartz, M. (2012). Fluency of speech depends on executive abilities: Evidence for two levels of conflict in speech production. *Procedia-Social and Behavioral Sciences*, *61*, 183-184. doi: 10.1016/j.sbspro.2012.10.138

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research, 31*(1), 117-134. doi: doi:10.1177/0267658314536435

Park, H., Rogalski, Y., Rodriguez, A. D., Zlatar, Z., Benjamin, M., Harnish, S., . . . Reilly, J. (2011). Perceptual cues used by listeners to discriminate fluent from nonfluent narrative discourse. *Aphasiology, 25(9),* 998-1015. doi: 10.1080/02687038.2011.570770

Pechmann, T., Garrett, M., & Zerbst, D. (2004). The time course of recovery for grammatical category information during lexical processing for syntactic construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 723.

Phillips, L. H. (1997). *Do ‘‘frontal tests’’ measure executive function? Issues of assessment and evidence from fluency tests.* In P. Rabbitt (Ed.), Methodology of frontal and executive function (pp. 191–213). Hove, UK: Psychology Press.

Postma, A., & Kolk, H. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research, 36(3),* 472–487. doi:10.1044/jshr.3603.472

Prins, D., Main, V., & Wampler, S. (1997). Lexicalization in Adults Who Stutter. *Journal of Speech, Language, and Hearing Research, 40*(2), 373-384. doi: 10.1044/jslhr.4002.373

Ratner, N. B., & Sih, C. C. (1987). Effects of Gradual Increases in Sentence Length and Complexity on Children's Disfluency. *Journal of Speech and Hearing Disorders, 52*(3), 278-287. doi: 10.1044/jshd.5203.278

Robinson, G., Blair, J., & Cipolotti, L. (1998). Dynamic aphasia: an inability to select between competing verbal responses? *Brain, 121*(1), 77-89. doi: 10.1093/brain/121.1.77

Robinson, G., Shallice, T., Bozzali, M., & Cipolotti, L. (2010). Conceptual proposition selection and the LIFG: neuropsychological evidence from a focal frontal group. *Neuropsychologia, 48(6),* 1652-1663. doi: 10.1016/j.neuropsychologia.2010.02.010

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and language*, *37*(3), 440-479. doi: 10.1016/0093-934x(89)90030-8

Salis, C., & Edwards, S. (2004). Adaptation theory and non-fluent aphasia in English. *Aphasiology, 18*(2), 1103-1120. doi: 10.1080/02687030444000552

Sampson, M., & Faroqi-Shah, Y. (2011). Investigation of Self-Monitoring in Fluent Aphasia with Jargon. *Aphasiology, 25*, 505-515. doi: doi:10.1080/02687038.2010.523471

Sanchez, G. (2013). PLS Path Modeling with R Trowchez Editions. Berkeley. http://www.gastonsanchez.com/PLS Path Modeling with R.pdf”

Sasisekaran, J., De Nil, L. F., Smyth, R., & Johnson, C. (2006). Phonological encoding in the silent speech of persons who stutter. *Journal of Fluency Disorders, 31*, 1-21. doi:10.1016/j.jfludis.2005.11.005

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, *54*(2), 199-227. doi: 10.1016/j.jml.2005.10.002

Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences*, *106*(1), 322-327. doi: 10.1073/pnas.0805874106

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of memory and language*, *29*(1), 86-102. doi: 10.1016/0749-596x(90)90011-n

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and language*, *54*(2), 228-264.

Schwartz, M. F., Saffran, E. M., & Marin, O. S. (1980). The word order problem in agrammatism: I. Comprehension. *Brain and language*, *10*(2), 249-262. doi: 10.1016/0093-934x(80)90055-3

Segars, A. H. (1997). Assessing the unidimensionality of measurement: A paradigm and illustration within the context of information systems research. *Omega*, *25*(1), 107-121. doi:10.1016/s0305-0483(96)00051-5 Warrington

Shao, Z., Meyer, A. S., & Roelofs, A. (2013). Selective and nonselective inhibition of competitors in picture naming. *Memory & cognition*, *41*(8), 1200-1211. doi: 10.3758/s13421-013-0332-7

Shao, Z., Roelofs, A., & Meyer, A. S. (2012). Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *The Quarterly Journal of Experimental Psychology*, *65*(10), 1927-1944. doi: 10.1037/e512592013-461

Silverman, S. W., & Ratner, N. B. (1997). Syntactic Complexity, Fluency, and Accuracy of Sentence Imitation in Adolescents. *Journal of Speech, Language, and Hearing Research, 40*(1), 95-106. doi: 10.1044/jslhr.4001.95

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In: Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008).

Spieler, D. H., & Griffin, Z. M. (2006). The influence of age on the time course of word preparation in multiword utterances. *Language and Cognitive Processes, 21*(1-3), 291-321. doi: 10.1080/01690960400002133

Sung, J. E., McNeil, M. R., Pratt, S. R., Dickey, M. W., Hula, W. D., Szuminsky, N. J., & Doyle, P. J. (2009). Verbal working memory and its relationship to sentence-level reading and listening comprehension in persons with aphasia. *Aphasiology, 23*(7-8), 1040-1052. doi: 10.1080/02687030802592884

Thompson, C. K., Cho, S., Hsu, C. J., Wieneke, C., Rademaker, A., Weitner, B. B., ... & Weintraub, S. (2012). Dissociations between fluency and agrammatism in primary progressive aphasia. *Aphasiology*, *26*(1), 20-43. 10.1080/02687038.2011.584691

Thorne, J., & Faroqi-Shah, Y. (2016). Verb Production in Aphasia: Testing the Division of Labor between Syntax and Semantics. *Seminars in Speech and Language, 37*(01), 023-033. doi: 10.1055/s-0036-1571356

Tompkins, C. A., Bloise, C. G., Timko, M. L., & Baumgaertner, A. (1994). Working memory and inference revision in brain-damaged and normally aging adults. *Journal of Speech, Language, and Hearing Research*, *37*(4), 896-912. doi: 10.1044/jshr.3704.896

Trupe, E. H. (Ed.). (1984). Reliability of rating spontaneous speech in the Western Aphasia Battery: Implications for classification: *Clinical Aphasiology Conference Proceedings.* pp.55-69. BRK Publishers.

Tsai, P. T., & Ratner, N. B. (2016). Involvement of the Central Cognitive Mechanism in Word Production in Adults Who Stutter. *Journal of Speech, Language, and Hearing Research, 59*(6), 1269-1282. doi: 10.1044/2016_JSLHR-S-14-0224

Varley, R., & Whiteside, S. P. (2001). What is the underlying impairment in acquired apraxia of speech. *Aphasiology*, *15*(1), 39-49.

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience & Biobehavioral Reviews*, *35*(3), 407-426. doi:10.1016/j.neubiorev.2010.04.007

Vigliocco, G., Vinson, D. P., & Siri, S. (2005). Semantic similarity and grammatical class in naming actions. *Cognition*, *94*(3), 91-100.

Wagenaar, E., Snow, C., & Prins, R. (1975). Spontaneous speech of aphasic patients: A psycholinguistic analysis. *Brain and Language, 2*, 281-303. doi: http://dx.doi:10.1016/S0093-934X(75)80071-X

Wang, J., Marchina, S., Norton, A. C., Wan, C. Y., & Schlaug, G. (2013). Predicting speech fluency and naming abilities in aphasic patients. *Frontiers in Human Neuroscience, 7*, 831. doi: 10.3389/fnhum.2013.00831

Wertz, R. T., Deal, J. L., & Robinson, A. J. (1984). *Classifying the aphasias: A comparison of the Boston Diagnostic Aphasia Examination and the Western Aphasia Battery*. Paper presented at the Clinical Aphasiology Conference, Seabrook Island, SC. http://aphasiology.pitt.edu/794/

Wilshire, C. E., Coslett, H. B., Nadeau, S., Gonzalez Rothi, L. J., & Crosson, B. (2000). Disorders of word retrieval in aphasia: Theories and potential applications. In S.E. Nadeau, L. J. G. Rothi, & B. Crosson (Eds). Aphasia and language: Theory to practice, pp. 82-107. New York: Guilford Press.

Wingate, M. E. (1988). *The structure of stuttering*. New York: Springer Verlag.

Ziegler, W. (2011). Apraxic failure and the hierarchical structure of speech motor plans: a non-linear probabilistic model. In A. Lowit R.D. Kent & (Eds.), *Assessment of Motor Speech Disorder* (pp. 305-323). San Diego: Plural Publishing Group.

Zingeser, L. B., & Berndt, R. S. (1990). Retrieval of nouns and verbs in agrammatism and anomia. *Brain and language*, *39*(1), 14-32. doi:10.1016/0093-934x(90)90002-x

Appendix A

Fluency scoring system of the Western Aphasia Battery -Revised (Kertesz, 2006) WAB-R

0 = No words or short meaningless utterances.

1 = Recurrent, brief, stereotypic utterances with varied intonation; the emphasis on prosody may convey some meaning.

2 = Single words, often paraphasia, effortful and hesitant.

3 = Longer, recurrent stereotypic or automatic utterances without information, or mumbling.

4 = Halting, telegraphic speech; mostly using single words; paraphasias; occasional prepositional phrases; severe word-finding difficulty. No more than two complete sentences with the exception of automatic sentences (e.g., “Oh I don’t know.”); characteristic of agrammatic nonfluent aphasia.

5 = Often telegraphic but more fluent speech with some grammatical organization; marked word-finding difficulty. Paraphasias may be prominent; few but more than two propositional sentences.

6 = More propositional sentences with normal syntactic patterns; may have paraphasias; significant word-finding difficulty and hesitations may be present.

7 = Phonemic jargon with semblance to English syntax and rhythm with varied phonemes and neologisms. May talk excessively; must be fluent; characteristic of severe Wernicke’s aphasia.

8 = Circumlocutory, fluent speech; moderate word-finding difficulty; with or without paraphasias; may have semantic jargon. The sentences are often complete but may be irrelevant.

9 = Mostly complete, relevant sentences; occasional hesitations and/or paraphasias; some word-finding difficulty; near normal but still perceptibly aphasic.

10 = Sentences of normal length and complexity, without definite slowing, halting or paraphasias.

Appendix B

Additional pls-pm results

Table B1 - Loading and cross-loading of manifest variables under each latent variable on that and other latent variables in the model. See Table 2 for information on each measurement.

Manifest Variable	Latent Variable	comp	word	wm	restart	syntax	fluency
WAByesno	comp	0.75	0.39	0.35	-0.04	0.30	0.28
WABaudWrec	comp	0.83	0.52	0.51	0.07	0.40	0.45
WAB sentence comprehension	comp	0.83	0.55	0.51	0.03	0.26	0.25
WAB object naming	word	0.56	0.88	0.51	0.08	0.21	0.25
Verb Naming Test	word	0.54	0.88	0.63	-0.23	0.41	0.47
BonnieBoston Naming Test (revised scoring)	word	0.54	0.91	0.67	-0.04	0.28	0.42
AphasiaBank Repetition 1A	wm	0.37	0.57	0.81	-0.03	0.23	0.36
AphasiaBank Repetition 1B	wm	0.30	0.51	0.82	0.07	0.25	0.20
AphasiaBank Repetition 1B- Serial	wm	0.27	0.48	0.79	0.05	0.23	0.17
AphasiaBank Repetition 2A	wm	0.64	0.59	0.87	-0.04	0.48	0.46
AphasiaBank Repetition 2B	wm	0.63	0.66	0.90	-0.05	0.49	0.56
Retracing	restart	0.08	-0.05	-0.02	0.82	0.06	-0.02
Repetition	restart	0.03	-0.09	-0.01	0.99	-0.12	-0.13
% Verbs	syntax	0.11	0.10	0.16	-0.09	0.66	0.35
DSS	syntax	0.25	0.17	0.22	0.00	0.76	0.48
% Grammatical	syntax	0.37	0.49	0.45	-0.13	0.70	0.44
% Prepositions	syntax	0.35	0.18	0.38	-0.04	0.65	0.42
% Past tense verbs	syntax	0.18	0.08	0.13	-0.02	0.63	0.27
Words/Second	fluency	-0.04	-0.08	-0.03	-0.16	0.45	0.48
WAB spontaneous fluency score	fluency	0.46	0.49	0.51	-0.09	0.55	0.98

Table B2- Results of the inner model with fluency only represented by words/second, obtained by 10,000 bootstrapping iterations. Rows with all zeros represent connections dependencies that are not defined in the model.

Direct effects	**Original path coefficient**	**Mean bootsrapped coefficient**	**SE**	**95% CI - Lower bound**	**95% CI - Upper bound**
comp -> word	0.55	0.55	0.07	0.42	0.68
comp -> wm	0.21	0.22	0.08	0.06	0.38
word -> wm	0.56	0.56	0.08	0.40	0.71
wm -> syntax	0.41	0.42	0.07	0.27	0.56
comp -> fluency	-0.09	-0.08	0.09	-0.26	0.09
word -> fluency	-0.15	-0.14	0.15	-0.42	0.14
wm -> fluency	-0.14	-0.15	0.13	-0.39	0.10
restart -> fluency	-0.12	-0.12	0.07	-0.23	0.07
syntax -> fluency	0.61	0.61	0.07	0.47	0.74
Total effects (direct _ indirect effects)					
comp -> word	0.55	0.55	0.07	0.42	0.68
comp -> wm	0.51	0.53	0.06	0.40	0.64
comp -> syntax	0.21	0.22	0.05	0.13	0.33
word -> wm	0.56	0.56	0.08	0.40	0.71
word -> syntax	0.23	0.24	0.05	0.14	0.34
wm -> syntax	0.41	0.42	0.07	0.27	0.56
comp -> fluency	-0.11	-0.10	0.09	-0.28	0.08
word -> fluency	-0.09	-0.08	0.12	-0.30	0.17
wm -> fluency	0.11	0.11	0.14	-0.15	0.38
restart -> fluency	-0.12	-0.12	0.07	-0.23	0.07
syntax -> fluency	0.61	0.61	0.07	0.47	0.74

Figure captions

Figure 1- The architecture of the fluency pls-pm model. The circles show latent variables, and the rectangles the manifest variables (see also Table 2). The large gray circle represents the inner model. The direction of the arrows indicates which variable exerts an influence over the other variables. Note that the latent variables in the model are all reflective.

Figure 2- Box plots for the two measures of fluency in the participants (N = 112). The left panel shows fluency measured as words/second. Right panel shows fluency measured as WAB-R spontaneous fluency scores.

Figure 3 - Loading (diagonal cells) and cross-loading (off-diagonal cells) of the manifest variables under each latent variable on that variable and other latent variables in the model. The dashed line marks the criterion 0.6. The diagonal (black-outlined boxes) show that the tests that are meant to measure a certain variable indeed have the highest loading on that variable. See Table B1 in Appendix B for the values of loading and crossloading for each test.

Figure 4- Parameter estimates for the inner and outer models in the fluency pls-pm. The values on the arrows from the latent variables (circles) to manifest variables (rectangles) are factor loadings. The values on the arrows between the latent variables are path coefficients. The values inside the parentheses next to the path coefficients are their corresponding t statistics. See Table 3 for 95%CIs obtained through bootstrapping. The values inside the circles are DG rho's.

Figure 5 - Direct and indirect effects of the five independent variables on the dependent variable fluency. The full height of the bars shows the total effect.

Footnote

1- Bootstrapping is a technique for estimating the sampling distribution of a statistic using random sampling with replacement.