UNIVERSITY OF GOTHENBURG
DEPT OF LANGUAGES AND LITERATURES

THE INFLUENCE OF EDUCATIONAL LEVEL AND OCCUPATIONAL STATUS ON THE SPOKEN LANGUAGE PRODUCTION OF PERSONS WITH AGRAMMATIC APHASIA

Baran Johansson

Department of Languages and Literatures

University of Gothenburg

English: Master Thesis, Linguistic Specialization

23 December, 2012

Supervisors: Prof. Jennifer Herriman & Prof. Elisabeth Ahlsén

Examiner & Opponent: Dr. Rhonwen Bowen

Abstract

Can educational background and occupational status have an influence on the spoken language production of persons with agrammatic aphasia? This study is an attempt to answer this question based on the investigation of speech production of three American high school and three American university educated persons with agrammatic aphasia. Syntactic, morphological, semantic, phonological and lexical analyses have been performed on the data. Part of the syntactic and lexical analysis of this paper is compared with the corpus findings of the Longman Grammar of spoken and written English. The results of the analyses have been compared both within the participants and between the groups. The findings of this study show that there is a difference between the language performances of these two groups. The university graduate subjects used a greater number of words and grammatical categories and they made considerably less linguistic errors in their speech than the high school graduate participants.

Acknowledgement

iv

1 Introduction

Languages across the globe connect people together every day. "Conversation is the most basic form of human communication" (Biber et al. 1999, p.16). One of the main differences between conversation and other registers such as fiction, news and academic prose is that conversation is a face to face interaction. The interlocutors involved in the conversation not only share the same physical context of space and time but also a great amount of personal and social knowledge. The interlocutors also usually have a communicative purpose and they communicate about their personal lives and interests (Biber et al., 1999). One of the characteristics of conversation is that it is spontaneous. Therefore, the speakers do not have much time to plan ahead and the utterances take place in "real time" (Biber et al. 1999, p.1048). As a result, it is quite common for the speaker to use repetitions as in: *the - the*, hesitators such as: *er, um* and contractions as in: *it's, aren't*. Another characteristic of the spoken language is the frequent usage of ellipsis. Questions and imperative sentence types elicit a response. These two types of sentences are also much more common in spoken language than for example written language (Biber et al., 1999).

Analyzing spoken language is a difficult task as the speaker can fail to complete an utterance which results in grammatically incomplete utterances. Biber et al. (1999) point out four situations in a conversation where the speaker does not succeed to finish a grammatical unit: *self- repair*, *interruption*, *repair by another speaker* and *abandonment of the utterance* (Biber et al. 1999, p.1063). In the first situation, *self-repair*, the speaker disregards a piece of discourse and starts fresh. In the second situation, the speaker is *interrupted* by another speaker. In the third situation, *another speaker is cooperating* with the first speaker to complete the utterance and in the last situation, the speaker, without any interruption, totally *abandons the utterance*.

This basic type of communication is an issue for people who have suffered from some form of aphasia. Aphasia is an impairment or deficit in language function due to brain damage (Soares and Ortiz, 2008). According to the statistics, there are approximately one million persons with aphasia in the United States of America. (National Stroke Association 2012, p.2)

Language is an extremely complex system. Therefore, there are different types of impairment related to this system and as a result there are various types of aphasia. One type of acquired brain damage could cause Broca's aphasia. This disorder is usually a result of trauma or stroke

(Avrutin, 2001). "Broca's Aphasia is the most common of the nonfluent aphasias. It is termed after a French physician, Paul Broca, in 1865. The lesion that causes Broca's aphasia affects the third frontal convolution (both the gyrus and the sulcus) of the left frontal lobe. This location is called Broca's area" (McCaffrey, 1998). The Western Aphasia Battery (WAB) is an instrument for measuring the degree, and type of aphasia. WAB AQ (Aphasia Quotient) score demonstrates the severity of language impairment in the subjects (Wikipedia, 2012). According to Steele (2007, p.9), WAB AQ score from 8 to 32 for persons with Broca's aphasia demonstrates Severe Broca's aphasia. WAB AQ score from 32 to 56 shows Moderate Broca's aphasia and WAB AQ score from 56 to 80 displays Mild Broca's aphasia.

One of the most studied aspects of the classical syndromes of Broca's aphasia is Agrammatic aphasia. It has been given the most theoretical and experimental attention by different scholars in the past decades (Menn, Obler and Miceli, 1990). Typical characteristics of agrammatism are "slow, halting speech, short and/or fragmentary sentences, limited output use of the syntactic and morphological resources of language. For example in English, subjects tend to drop out articles, connective words, auxiliaries and inflections and leave the lexical words such as nouns, verbs and adjectives" (Albert et al. 1981, p.153).

The effect of education on speech language production of persons with agrammatic aphasia is an area where previous researchers differ in their conclusions; see section "previous research" for more background information. This gap in our understanding of the influence of educational level in the speech language production and the importance of spoken language as well as the fact that persons with agrammatic aphasia struggle with this phenomenon lead to the question this paper seeks to answer:

Considering a selection of linguistic analyses: syntactic, morphological, semantic, phonological and lexical, are there any differences between the spoken language production of high school and university educated persons with agrammatic aphasia?

1.1 Previous Research

1.1.1 The development of the research field: *Agrammatism*

One of the influential theories in agrammatic studies, *economy of effort hypothesis*, was introduced by Isserlin (1922). According to this researcher, speech production is a very exhausting process for persons with agrammatic aphasia, as articulating the words is a burden for them, therefore they try to adapt themselves to this situation by using as few words as they can which would result in producing simple structures. This theory was later developed. Kolk and Heeschen (1990) presented a theory called *adaptation theory*. They claimed that telegraphic speech of persons with agrammatism is the result of adaptation. In this case, the person tries to avoid any sentence production problem by using simple elliptical expressions. This is a strategic choice made by the subject.

The paradigms that Chomsky introduced in 1957 and further developed in 1965, so called transformational grammar, or generative grammar and later minimalism had a certain effect on aphasia research (Ahlsén, 2006). The abstract theory, transformational grammar that Chomsky (1957) presented implies that there is a "deep structure" and a "surface structure" in each sentence of a language. Chomsky also introduced the concept of "language competence" which was developed by other linguists. Chomsky makes a distinction between competence, which is "the speaker-hearer's knowledge of his language", and performance being "the actual use of language in concrete situations" (Chomsky 1965, p.4).

In 1971, Harry Whitaker was inspired by Chomsky's transformational grammar and claimed that some types of aphasia, for example agrammatism, are caused by a disorder of competence (Whitaker, 1971).

In 1980 Saffran and his colleagues introduced the *mapping theory*. They found that persons with agrammatism have problems putting the main semantic relations such as: *agent* and *action* into a correct word order in a sentence (Saffran, Schwartz and Marin, 1980).

Another significant theory in this field was proposed by Berndt and Caramazza (1980) which is recognized as *syntactic hypothesis*. According to this hypothesis, the occurrence of telegraphic speech and agrammatic comprehension is due to the fact that the person with Broca's aphasia is incapable of using syntactic information.

An extensive cross-linguistic study of persons with agrammatic aphasia, in 14 languages, has been conducted by different scholars which provides detailed grammatical descriptions and distributional analyses (Menn, Obler and Miceli, 1990).

Researchers who have studied agrammatism have mostly focused on one linguistic phenomenon. For example Bastiaanse and Thompson (2003) investigated verb and auxiliary movement in persons with agrammatic aphasia and found that for English speaking persons with agrammatic aphasia, sentence production is more impaired once the auxiliary verb is moved from its base generated position than when the finite lexical verb or the auxiliary is in its base-generated position. Another study carried out by Fix and Thompson (2006) showed that agrammatic production of irregulars is not dependent on their morphological structure. Therefore, all irregulars are processed similarly. In another investigation Burchert and his colleagues indicated that both comprehension and production of complex sentences are harder than simple sentences for persons with agrammatic aphasia (Burchert et al., 2007).

1.1.2 The effect of education on the subjects

There were also other studies which focused on the effect of different socio-demographic variables on the linguistic performance of persons with aphasia. For example, Béland and Lecours (1990) examined the difference between the language performances of two healthy adult groups. They discovered that school educated adults with longer education performed better on some different language tasks such as *verbal fluency, repetition of word* and *sentence picture-matching* than subjects with shorter education.

Tainturier, Tremblay and Lecours (1992) conducted a study in which they investigated the connection between educational length and the effect of word frequency. They chose 40 right-handed healthy adults. Half of the subjects studied less than 15 years (on average 11.4 years) and the other twenty participants had studied 15 years or more (on average 18.4 years). All the subjects were French native speakers and fluent readers. The majority of them knew some English as well. They discovered that it takes more time for high-school graduate subjects to recognize lower frequency words than for university graduates. They concluded that educational length of the participants should always be taken into account as a potentially dominant variable in the analysis of language-related experimental studies.

Connor et al. (2001) studied 39 subjects with aphasia both at about 4 months and 103 months postonset. They discovered that the subjects with lower educational levels and occupation

status had considerably higher degrees of aphasia. However, these two demographic variables did not play any role in the rate of recovery of the participants.

Lazer et al. (2008) investigated the nature and variability in language recovery of 22 persons with aphasia from 24-72 hours after stroke onset and they followed and checked their language performance and recovery till 90 days after the stroke. They claim that lesion size, age and educational background of the subjects do not have any effect on the language performance and recovery of the subjects at 90 days.

Soares and Ortiz (2008) examined the language assessment of thirty persons with aphasia who had different educational background. Half of the subjects had 1-4 years of education and the other half studied between 5-11 years at school. The researchers compared the language performance of the persons with aphasia with the control group. The control group were healthy subjects of the same age, sex and length of education. The researchers of this study found that the persons with aphasia failed to access the lexicon in the verbal fluency tasks and the higher educated subjects did not perform better than the other participants with aphasia. Their results did not show any difference between the language performances of these two groups of participants. However, there was a difference between the performances of the control group regarding their educational length. They concluded that the reason that higher educated subjects did not perform better in the language tasks is that what affects their performance is language impairment due to their brain injury and not their educational background.

Recently a study was carried out by Fernandez et al. (2011). They aimed to show the association of educational level and socioeconomic status on aphasia severity after stroke. They analyzed the error percentage of some language tasks such as *auditory* and *written comprehension*, *written naming*, *oral reading* and *spelling* of 173 participants within 24 hours after they had a stroke. They discovered that the subjects who had studied 12 or more years made considerably fewer errors than the other group with less than 12 years of education.

As presented in the section above, we can clearly see contradictory results about the importance of educational length. Some studies have found education as an important factor and some have not.

2 Design of the Present Study

This investigation focuses on the spoken language of six persons with non-fluent Broca's aphasia, who, in this study are referred to as persons with agrammatism (PWAgr) from now on.

Two different types of persons with agrammatism were selected: high school graduates and university educated persons. The three PWAgr in the high school graduate group had studied up to high school level. Only one of those three participants had studied one year at a college. All the participants in this group had physically challenging jobs before they had a stroke. The three PWAgr among university graduates have studied at least six years at university and obtained a Master's degree in one subject. The participants of this group had mentally challenging jobs before their stroke. A number of linguistic analyses were performed on the spoken language of these two groups. In general syntactic analysis performed on the data provides information on different forms of grammatical categories used correctly or incorrectly by the subjects as well as the mean length of the sentences produced by the participants. Morphological analysis would give us an insight into different inflectional morphemes that the participants used correctly and incorrectly in their interviews. Lexical analyses would show type token ratio and neologistic errors made by the participants of this study. According to Goodglass and Kaplan (1983, p.8) "Paraphasia refers to the production of unintended syllables, words or phrases during the effort to speak". As paraphasia is an important characteristic of speech production of PWAph, all three types of paraphasias: phonological, semantic and neologistic will be investigated in this paper.

2.1 Aim

The objective of this project is to investigate the spoken language of these two groups, high school graduate PWAgr and university graduate PWAgr in order to find out if there is a difference between their language production related to their educational backgrounds and the type of job they had before their illness. My hypothesis is that being exposed to formal language, academic writing and scientific articles subconsciously affects the person's language production. Generally, university graduates are more used to this type of language; therefore they automatically follow the linguistic rules and speak more correctly than high school graduates. However, it is not only about having a university degree. One should also have a job which demands the person to be updated and read these types of materials. One can have no university degree but the type of job that person has requires him to read scientific

articles, attend seminars and be familiar with formal language. Therefore these two variables, educational background and occupation, go hand in hand. Thus, there is a difference between the language of high school graduates who have physically challenging jobs and university graduates with mentally challenging jobs. The question is if this assumed difference would also affect these two groups' language production after their stroke. In order to answer this question the spoken language of these six subjects is analyzed from a number of linguistic perspectives. The usage of grammatical categories, total number of words produced, number of grammatical categories produced by each subject and each group, inflectional morphemes which were used or mistakenly not used by participants of each group. Type-token ratio, sentence length as well as different types of error analysis: grammatical, semantic, neologistic and phonological have been performed on the data. In addition, in order to have some valuable source to compare these findings with, the Longman Grammar of Spoken and Written English has been chosen. Therefore, some of the results of this study were compared with the Longman's conversation corpus findings. This comparison was necessary in order to grasp how far or close these participants have performed in relation to people in general.

It is also important to point out that five of these subjects are Mild PWAgr. There is one participant, Scale 1a, in the university graduate group who is Moderate PWAgr. The reason for choosing one Moderate in this group is to find out whether this participant who has a more severe language impairment would perform better than Mild high school graduate PWAgr in the tasks. In case, he does, it would only make my hypothesis stronger.

2.2 Material

The primary material in this study consists of transcriptions and videos of six PWAgr speech during some interviews conducted by the clinicians of www.talkbank.org (Talkbank, 2012a). The Talkbank system provides researchers with a data base of multimedia data on human communication for studying a variety of language phenomena such as child language development (childes.psy.cmu.edu), second language learning (talkbank.org/BilingBank) and aphasia (MacWhinney et al., 2010).

The subjects of this study are chosen from AphasiaBank (Talkbank, 2012b;f). AphasiaBank consists of a computerized database of interviews between PWAph and clinicians. In order to collect these interviews, the clinicians use a consistent protocol format (MacWhinney at al., 2011). This protocol includes different tasks such as: two free speech elicitations, four picture descriptions, one story telling (Cinderella) and finally one procedural discourse task

(MacWhinney et al., 2010). These tasks will be discussed more in detail in the language data collection section.

2.2.1 Subjects

2.2.1.1 Background Information of the Subjects

In order to choose the participants for this study, I have considered some social variables such as: *age, gender*, *occupation*, *educational background* as well as some neurolinguistic variables, for example all the subjects were required to be categorized as persons with non-fluent Broca's aphasia by clinical standards. In addition, some neurological variables such as: *handedness*, *etiology*, *general health*, *aphasia duration, WAB AQ* scores and linguistic variables such as *language status*, *primary and other languages learned by the subjects* were also taken into account in selecting the subjects.

Table 1 and 2 represent the background information of each subject in more detail. Talkbank has presented the university graduates used in this study as: Adler 4a, Scale 1a, Thompson 3a and the high school graduate participants as Tap 14a, Scale 5a and Scale 18a. However, in order to increase readability for the reader, the subjects will be renamed from now on as follow: Adler 4a (U1), Scale 1a (U2), Thompson 3a (U3), Tap 14a (H1), Scale 5a (H2) and Scale 18a (H3). The "U" stands for "university graduate subjects" and "H" stands for "high school graduate participants".

Table 1. Background Information of the University Graduates

	University Graduate Participants		
	Adler 4a (U1)	Scale 1a (U2)	Thompson 3a (U3)
Age at Testing	75.5	78.3	67.6
Gender	Female	Male	Male
Profession	Professor	President of T. (CEO)	CPA
Years of Education	20	18	18
Aphasia Category-- Clin Impression	NFL	NFL	NFL
Aphasia Type--Clin Impression	BRO	BRO	BRO

WAB AQ. (max = 100)	72.6	52.5	72.6
Handedness	Right	Right	Right
Aphasia Etiology	Stroke	Stroke	Stroke
General Health	Good	Excellent	No complaints
Language Status	Multilingual	Monolingual	Monolingual
Primary Languages	English	English	English
Other Languages, in order learned	French, 2 Middle Eastern	-	-
Aphasia Duration (years)	6.0	25.09	14.03

Subject 1. U1

U1 was a university professor. She had a stroke after heart surgery six years before her interview. This subject was categorized as having non-fluent Broca's aphasia and according to (Steele, 2007) U1's WAB AQ score demonstrates that she suffers from a Mild Broca's aphasia. She was 75.5 years old at testing. She had studied eight years at university. Therefore, she has the longest education among the subjects of this study. This participant is an American native speaker and according to Talkbank, she was a multilingual and knew French and two Middle Eastern languages. However, there is no further information about which languages these two Middle Eastern languages might be and how her performance has been in these three languages after her illness. She is fully right-handed.

Subject 2. U2

U2 was CEO of a company before his stroke. He was clinically categorized as having non-fluent Broca's aphasia and as his WAB AQ score shows his brain damage is more severe than the other subjects in this study and as a result is the only Moderate PWAgr in this investigation. He got a stroke 25.09 years before his interview on Talkbank. His age at testing was 78.3 which makes him the oldest subject in the present study. However, according to Talkbank he was in excellent health when the interview was conducted. He is a monolingual American native speaker who has studied six years at university. He is also right-handed.

Subject 3.U3

U3 worked as CPA (Certified Public Accountant) prior to his stroke which happened 14.03 years before the interview. He was 67.6 years old at the time of testing. This participant was categorized as having non-fluent Broca's aphasia. His WAB AQ score indicates that he suffers from Mild Broca's aphasia. There were no complaints from the interviewer about this participant's general health. U3 studied six years at university and obtained a Master's degree in accounting. After listening to his speech, one would understand that he has worked with numbers before his stroke as he pays special attention to all the dates and figures as well as names of the restaurants he used to dine in and more details. His mother tongue is American English and there is no information indicating his proficiency in other languages. He is a right-handed.

Table 2. Background Information of the High School Graduates

	High School Graduate Participants		
	Tap 14a (H1)	Scale 5a (H2)	Scale 18a (H3)
Age at Testing	44.9	63.7	49.6
Gender	Male	Male	Female
Occupation	Fuel tanker driver	Tour bus driver	Paratrooper
Years of Education	12	13	12
Aphasia Category-- Clin Impression	NFL	NFL	NFL
Aphasia Type--Clin Impression	BRO	BRO	BRO
WAB AQ. (max = 100)	60.2	73.2	60.9
Handedness	Right	Right	Left
Aphasia Etiology	Stroke	Stroke	Stroke
General Health	Excellent	Good	History of hypertension
Language Status	Monolingual	Monolingual	Monolingual
Primary Languages	English	English	English

Other Languages, in order learned	-	-	-
Aphasia Duration (years)	1.03	5.08	15.02

Subject 4.H1

H1 was a fuel tanker driver. His age at testing was 44.9 and he is the youngest participant in this paper, and he was 1.03 year post-stroke. He suffers from Mild non-fluent Broca's aphasia after his stroke. At the time of interview, he was in excellent health condition. He is a right-handed monolingual American native speaker and he has studied up to high school level.

Subject 5. H2

H2 used to work as a tour bus driver before his stroke. He was categorized as Mild non-fluent Broca's aphasia 5.08 years prior to this interview. He was 63.7 years old at testing. He is right- handed and was in a good health condition at the time of recording the video. It is not clear if he has studied one year at college or university before. He is a monolingual American native speaker.

Subject 6. H3

Considering the duration of aphasia among all the subjects, H3 was the youngest when she got a stroke 15.02 years before the testing. She was 49.6 years old at the time of interview. She mentions that she worked as a paratrooper. She was recognized as Mild Broca's aphasia and is the only left-handed subject in this paper. It is mentioned in the general health condition section on Talkbank that this subject has had a history of hypertension. She is a monolingual American native speaker.

2.2.2 Data Collection

The goal of this paper was to investigate the language of six PWAgr within two groups with different educational length and careers. In order to achieve this aim, I went through all the participants' demographic database and tried to find those categorized as non-fluent Broca. Further on, I tried to narrow down my options by considering the following variables: *age, gender*, *occupation*, *educational background, handedness*, *etiology*, *general health*, *aphasia duration, WAB AQ* scores and *language status*. Nearly 40-60% of the persons with aphasia move from the acute phase to the non-acute phase within 6 to 12 months after their stroke.

This stage is characterized as chronic phase where the condition persists (Pedersen et al., 2004). Hence, it is important to take the *aphasia duration* variable into account in order to make sure that the participants were not in the acute phase while recording. However the difference in time post onset does not make a difference and the status at the time of the recording is the most important variable. According to Knecht and his colleagues, "in most people the left hemisphere of the brain is dominant for language." (Knecht et al. 2000, p.2512). They found that right-hemisphere language dominance in 100% right-handers was 4% comparing to ambidextrous persons to 15% and it increases to 27% in 100% left-handers (Knecht et al., 2000). That is the reason I have considered the variable *handedness* in this study.

All the persons with agrammatic aphasia, in this study were adults (above 18 years old) both at the onset of their stroke and at the time of testing. In addition, they are all native speakers of American English. All of the participants except H3 are right-handed. U1 is the only multilingual subject in this study and the rest are monolinguals. The major distinction between these two groups is the type of occupation and the educational length they have. The university graduate participants have studied at least six years at university and had mentally challenging jobs before their stroke. On the other hand, the high school participants had studied until high school level. Only one of them, H2, studied at college for one year. All of the subjects in this group had more physically challenging jobs before their illness.

Going through the AphasiaBank database, I discovered that there are more university educated PWagr in this site than high school graduates and not all of them talk so much. In order to capture a good understanding of these subjects' language ability to perform different linguistic analysis on the data, I decided to choose PWAgr who produced at least 250 words. In the high school graduate group the choices were very limited. There were only three Mild PWAgrs who used at least 250 words.

There were more options for university graduate participants. However, I tried to keep the participants' WAB AQ score as close as possible to one another in both of the groups. One subject among university graduates was chosen from Moderate PWAgr on purpose. My intention for choosing this person with Moderate Broca's aphasia, U2, was to test my hypothesis even further. I would like to know if the subject with Moderate Broca's aphasia from the university graduate group would perform better on the tasks than the Mild Broca subjects in the high school graduate group.

Having mentally jobs like the university graduate subjects demands a lot of reading textbooks and even scientific articles in order to update oneself and being able to encounter and fix the problems in the company as a CEO and CPA. U1 used to work as a university professor and that also requires doing research. Therefore, I assume that these three PWAgr were exposed to more standard and formal language than the high school graduate participants. Unfortunately there is no information on Talkbank about these subjects' interests and hobbies. However, we do know that working as a paratrooper and driver are more physically challenging jobs and do not require reading the type of texts mentioned above in order to perform better in their working environment.

2.2.3 Language Data Collection

The subjects were requested to talk about five different topics during interviews conducted by Talkbank researchers (Talkbank, 2012d):

2.2.3.1 Tasks

I. Free speech samples:

Stroke Story and Coping - In this task, the participants were required to talk about the history of their illness. There were also some follow-up questions from the interviewer, e.g. an important event in their lives (no matter if it was a sad or happy event which happened recently or in the past)

II. Description of four pictures (Talkbank, 2012e):
- Refused Umbrella
- Broken Window
- Cat Rescue
- Flood

III. Story Narrative

Cinderella - This fairy tale was included in AphasiaBank protocol because people from Western cultures are familiar with it (MacWhinney et al. 2010, p.857). The subjects were given a book full of pictures from this fairy tale and they had time to look at the pictures. After a couple of minutes the interviewer took away the book and they could start narrating this story from what they had seen in the book and/or from what they generally remembered from the Cinderella story.

IV.   Procedural Discourse

In this task the participants were requested to describe how they make a peanut butter and jelly sandwich.

2.2.3.2 Transcriptions

As I started listening to the subjects and going through the transcriptions, I realized that Talkbank has not transcribed everything the participants said. For example, when the participants look at the pictures from Cinderella story, they remember some parts and they comment on some of the pictures. However, Talkbank, for some unknown reason, has not included those words in their transcriptions. In addition, a couple of other words were left out untranscribed from each video. Therefore, I listened and watched the videos and transcribed whatever Talkbank had missed in their transcriptions. In addition, every word from the transcription was categorized according to grammatical categories used in Longman Grammar of Spoken and Written English as I wanted to compare some of the findings from this investigation with Longman's. Another reason why I disregarded Talkbank's grammatical analyses was that some of Talkbank's grammatical categories appeared incorrect. For example, in this excerpt from U1's corpus "*well first in+house and then uh out um patient um but not in the regular*" the word *patient* is categorized as an adjective. However, if one both listens and watches the video, one would understand that she meant "*she was going to a place as a patient but not on the regular basis*". Therefore, "*patient*" is a noun in this example and not an adjective.

A small passage from U1, in the university graduate group, and H2, from the high school graduate group, is presented in the appendix. In addition, in order to enable the reader to understand how the spoken language performance of persons with agrammatism looks like comparing to people in general, I have also included a small passage from LGSWE spoken corpus in the appendix.

2.3 Method

2.3.1 General overview of the research methods in this investigation

The combination of instructions on Talkbank (2012b) and the article "*AphasiaBank: Methods for studying discourse*" have given me a great understanding of what kind of protocols Talkbank clinicians have used as a source of their interviews, procedure of interviews, follow up questions as well as the abbreviation Talkbank researchers have used in their transcriptions

and analyses. Some of the methods used by Menn, Obler and Miceli (1990) in *Agrammatic Aphasia -A Cross-Language Narrative Sourcebook* have also been used in this paper as this sourcebook has focused on the same category of persons with agrammatism and they had similar type of tasks in their interviews.

The *Longman Grammar of Spoken and Written English* (Biber et al., 1999) has been used as a source for all word class, morphological and lexical analyses. This grammar is one of the best sources for the analysis of grammar and spoken language. Some of the results from this research have been compared with Longman's conversation corpus findings. The reason I have chosen to compare the results is to find out the difference as in how close or far away these participants have performed comparing to people in general.

Another thing to point out is that the subjects might have repeated one word a couple of times as in "*and, and, and um and I go*" however, in this case the word "*and*" before "*um*" is counted only once in the analysis. The reason is that the subject's word finding is impaired so he tries to find the next word and repeating these words does not contribute something new. The reader should bear in mind that the aim of this study is not to investigate which grammatical categories each subject struggles with but to find out the difference between the language production of these two groups, university and high school graduates, considering the relative proportions of grammatical categories they have produced and to compare these distributions with Longman's. The most and least frequent categories used by each participant are compared both within the groups and between them. Had I counted "*and*" three times in the above example this would have given the wrong impression about the categories this subject used and could not have been comparable with Longman's findings. Therefore, what I consider as the target word has been used for the analysis for each subject. On the other hand, if the participant wanted to emphasize something and therefore has repeated a word a couple of times as when U1 said "*Cinderella is always working, working, working*". "*Working*" in this instance has been counted three times as she is intensifying and that is something that even ordinary people can use and it does not have anything to do with word finding problems.

In cases where the participant has used a grammatical category incorrectly or has omitted one word from his/her speech it has been taken into account. The sentences produced by the subjects are reconstructed in order to find these errors and omissions. However, if the reconstruction was not based on obvious linguistic errors but on personal guesses, it has been excluded from the analysis of "omission and/or incorrect" part. For example if the subject said

"*I have book*", one can think that an indefinite article "*a*" is missing from this sentence. However, if the subject only said "*dresses ball*" it would have been difficult to know what the subject actually wanted to say. There can be a lot of different interpretations regarding grammatical categories. Therefore, in such cases I have excluded this fragment from the "omission and incorrect" analysis. This part of analyzing grammar of persons with agrammatism is very difficult.

Morphological, semantic, neologistic and phonological errors produced by each subject have been analyzed according to Talkbank coding errors (Talkbank, 2012c). Some of the errors the participants made were not recognized by Talkbank researchers. For example in the following case made by H2 "*the brover [: brother] is going to get the kid [: cat]*". I think that "*brover*" is a phonological error here. However, it is not mentioned by Talkbank researchers. In some other examples, I have categorized the error as another type of, for example, semantic or morphological error than the one provided by Talkbank.

Wordsmith 5.0 (Lexical Analysis Software Ltd., 2012) has been used as an appropriate tool for computing type token ratio (TTR) and mean sentence length of the present paper. TTR results of this investigation are compared with Longman's TTR in conversation.

2.3.2 Longman Grammar of Spoken and Written English

The *Longman Grammar of Spoken and Written English (LGSWE)* describes the actual usage of grammatical features in four different registers: conversation, newspaper language, fiction and academic prose. Longman has used a corpus-based approach which entails that the grammatical descriptions are based on usage and patterns of structure in a large electronic collection of both written and spoken texts. The LGSWE corpus contains over 40 million words (Biber et al., 1999). There are 3.436 texts (3.929.500 words) from British conversations and 329 texts (2.480.800 words) from American conversations making 6.410.300 words in the spoken corpus (Biber et al., 1999). The LGSWE conversation corpus represents contemporary English as all these texts were produced after 1980 (Biber et al., 1999). A one-million word corpus corresponds to 140-150 hours of conversational interaction (Biber et al. 1999, p.27). Unlike previous conversational corpora, the LGSWE conversation corpus has been collected in a genuinely natural setting (Biber et al., 1999). A set of informants were used to represent the range of British and American English speakers across sex, age, social groups and regional spread. Around 75% of the participants were over 18 years old. These participants carried a high-quality tape recorder with them for one week and tape recorded all the

conversational interactions they had in that period of time. Afterwards, these conversations were transcribed orthographically for grammatical and lexicographic research (Biber et al. 1999, p.29).

One of the most important usages of LGSWE corpus-based study is to provide information about frequency of use. The LGSWE has normalized all frequency counts reported in the grammar section to a common basis per million words of text. Therefore making it easier to compare the results for different features. In some cases the quantitative findings in the grammar are presented as percentages (Biber et al., 1999).

In Longman, the corpus findings for different word classes in spoken language are only presented in different graphs as a comparison to other registers such as news, fiction and academic language. So, they are not presented in separate tables. Therefore, the only way to get the value of each word class in the spoken language was to calculate the relative proportion (RP) from the relative frequency (RF).

2.3.3 Syntactic Analysis

The distribution of grammatical categories in the transcriptions for each subject is presented in the results.

Four different types of information are demonstrated in the tables about grammatical categories: the first column presents different word classes and the second column "provided" shows the number and relative proportion of the words in each word class provided by that subject. The third column "omission" exhibits number, relative proportion as well as the percentage of each word class that is omitted by the subject. In order to find out the percentage of words omitted from each word class, the numbers of words provided plus omitted from each word class have been calculated. The fourth column "Incorrect" exhibits number, relative proportion of the incorrect usage of each word class by the subject (if any). In addition, the tables provide information about the percentage of the incorrect usage of each word class in each subject's corpus.

In case one word class such as quantifiers or *wh*-words is used as adverbs in the material, it is shown separately in the table. However, it is discussed under the category "adv" and is accumulated together to present the total number of adverbs the participant used in his speech.

Different forms of some of the grammatical categories such as pronouns (demonstratve, indefinite, nominative personal, accusative, interrogative, etc) are presented separately in the

tables. There are two reasons for this type of detailed information: first, one of the analyses performed on the data is to find out how many different grammatical categories the participants have produced and compare them with one another. Therefore, I tried to be as specific as I could and it did not suffice to just have all different forms under one category. Second, in case future researchers are interested in a particular category or want to have more details for each word class, it would make it easier for them to use this material. However, in order to make the text presented after the tables not so repetitive, I have included the detailed explanation of the tables 3-8 in the appendix section. The most interesting aspects from each table as the five most and least frequent categories as well as some extra comments (if any) are presented under each table.

The five most and least frequent grammatical categories of all the subjects in this study are presented together in Table 9 (p.38). All the grammatical categories are ranked from the most frequent to the least frequent for each participant. In order to make it easier for the reader to understand which categories are among these five and would be clear for the comparison within the groups and between them, I have highlighted the top five categories with grey and the bottom five with black.

The mean length of the sentences in each subject's corpus has been calculated by WordSmith tools and the results are compared with other subjects and between these two study groups. The number of words produced by each participant and each group has also been counted.

Longman Grammar of Spoken and Written English (Biber et al. 1999, p.55) groups the words into three different classes: lexical (content) words, functions and inserts. Lexical words contain four main word classes: nouns, verbs, adjectives and adverbs (Biber et al. 1999, p.62). Function words are: determiners, pronouns, numerals, prepositions, primary auxiliaries, modal auxiliaries, adverbial particles, coordinating conjunctions, subordinating conjunctions, *wh*-words, the negator *not*, existential *there* and the infinitive marker *to* (Biber et al. 1999, p.69). Typical examples of inserts are: interjections, greetings, farewells, discourse markers, attention signals, response elicitors, responses, hesitators, thanks, the politeness marker *please*, apologies and expletives (Biber et al. 1999, pp.93-94). The number of function words, lexical words and inserts are calculated for each subject. The LGSWE shows that the word classes are not distributed evenly across different registers (Biber et al. 1999, p.11). The distribution of the word classes of this study from each subject within and between the groups is compared to the distribution of these three word classes in Longman's conversation corpus.

2.3.4 Morphological Analysis

The distribution of different forms of inflectional morphemes is shown in the results. The categories studied in this paper that are marked by inflection are as follows: (Biber et al. 1999, p.57)

Nouns: a distinction has been made between noun plural regulars (*books*) and noun plural irregulars (*women*), possessive marker (*boy's*).

Verbs: third person singular present indicative (*lives, writes*) regular past tense marker (*lived*) irregular past tense marker (*wrote*) and *ing* marker (*living, writing*).

Different types of morphological errors made by the subjects are presented in the analysis. Talkbank has some error codes (Talkbank, 2012c) for morphological errors which are used in this paper and will be explained as follow:

[*m: 0s] = inflectional suffix *-s* is missing from the regular and/or irregular plural nouns.

[*m:0es] = third person singular present morpheme *-s* is missing.

[*m:0ed] = past tense *-ed* marker (in the regular and/or irregular examples) is missing.

[*m:0ing] = *-ing* marker is missing from the verb.

[*m:+ing] = superfluous progressive. For example simple present form "*go*" was required. However, the participant said "*going*" instead.

[*m:a] = verb agreement error. For example "*she have [*m:a] two cups*".

[*m:a: +s] = superfluous plural on nouns as in "*one books*".

2.3.5 Semantic analysis

Four different kinds of semantic errors were produced by subjects of this study:

[*s:r] = in these cases the error is a recognizable word in English and it is semantically related to the target word, as in "*spoon*" for "*fork*" *spoon [:: fork]*.

[* s:per] = the repetition of the word is no longer appropriate. For example, *The boy kicked the ball through the ball [:: window] [* s:per]* .

[* s:ur] = the error is a real word for which the target is known. However it is not semantically related to the target, and does not meet the criteria for phonological errors either. For example *hi [:: time]*.

[* s:uk] = the error is a real word in this case, however the target word of this word is not known as in, *I go wolf [*s:uk]*.

2.3.6 Phonological analysis

Three phonological errors made by the participants were included. The coding phonological error will be explained here:

[* p:w] = the error is a real word as in *heat [:: eat]*.

[* p:n] = the error is not a real word. In this case Talkbank has transcribed the error using IPA and attached @u to the error. For example *lεθ@u [: left]*.

[* p:m] = the error involves metathesis[1]. Talkbank has transcribed the error using IPA and attached @u to the error. For instance, *mɪdwεts@u [: midwest]*.

2.3.7 Lexical Analysis

2.3.7.1 WordSmith Findings

As mentioned earlier, for this type of analysis a tool called WordSmith has been used. WordSmith tool is an integrated suite of programs which allows researchers to look at how words behave in texts (Lexical Analysis Software Ltd., 2012). Two types of analyses, the type token ratio and the mean sentence length computation, were performed on the texts using this program. The mean sentence length is related to the syntactic analysis and is discussed in that section. However, type token ratio analysis is a part of lexical analysis. Therefore, it is discussed in the lexical analysis section.

The Type Token Ratio (TTR) of each subject's corpus is computed and the results are compared within the subjects of one group and then between these two groups (university graduates and high school graduates). Afterwards, LGSWE's conversation corpus type token ratio is compared with the TTR of the participants in this study.

---

[1] A change in the order of sounds or letters in a word (Oxford University Press, 2011).

2.3.7.2 Analysis of Neologisms

Two different types of neologistic errors were produced by the subjects. These two kinds of errors will be discussed below:

[* n:k] = the error is a non-word and the target is known. However, it does not meet criteria for phonemic or semantic errors. Talkbank has transcribed the error using IPA and attached @*u* to the error. For instance, *gɹæstɪdʒɪz@u [: groceries]*.

[* n:uk] = the error is a non-word and the target is unknown. Talkbank has transcribed the error using IPA and attached @*u* to the error. In addition, Talkbank researchers have used this symbol [: x@n] as the target word. For example, *two ɹɛsɪz@u [: x@n]*.

2.3.7.3 Some Notes

The proper nouns such as *Wall-Street-Journal* are counted as one word.

In case the subject said something such as *I wanna* this is counted as three words *I want to*.

The contractions such as *I've* are counted as two words *I have*.

2.4 Limitations

Handedness can have an effect on the localization of pathology (www.Your-Neurologist.com, 2012). Therefore, all the participants were supposed to be right-handed. However, as I did not have many options to choose the subjects from, one of the participants, namely H3, is left-handed.

In the beginning of this study, the aim was to find subjects who were over 18 and under 65. However, as there were not many subjects to choose from on Talkbank, as a result two of the participants, U2 and U3, were over 65 at the time of testing and one of them, U1, was over 65 both at the time of testing and at onset. Yet all these subjects belong to the university-educated group. The goal was to study only monolingual subjects. However, as there were very few subjects available one of the participants, U1, is multilingual.

Finally, discourse analysis as well as different types of syntactic analysis such as: study of phrases, clauses and sentence are not within the scope of this investigation due to time limitations and not because they are of less interest and importance.

3 Results

In this section, first the findings for each subject will be presented independently. Syntactical, morphological, semantic, lexical and phonological analyses have been performed on the data and will be shown separately. At the end of this section, the interpretation of the findings will be discussed.

3.1 Raw Data Analysis

3.1.1 Syntactic Analysis

As mentioned in the method section, the syntactic analysis is based on the definitions and examples provided by *Longman Grammar of Spoken and Written English* (Biber et al., 1999).

3.1.1.1 Distribution of Grammatical Categories within subjects

The following six tables show different forms of grammatical categories used by each participant followed by the total number of each category produced by that subject and the relative distribution of those categories as percentage of the total number of words. In addition, the table demonstrates if the subject has omitted some of the required words or has used them incorrectly. The total number of omitted words and the total number of incorrect words in each category (if any) are shown. Moreover the relative proportions and percentage of the total number of provided words as well as omission for the omission part and percentage of provided words for the incorrect word classes are also presented.

Subject 1. U1

Table 3. Distribution of word classes

Word classes	Provided		Omission			Incorrect		
	N	RP %	N	RP %	% of total words (provided+omission) in word class	N	RP %	% of provided in word class
Adjectives	31	3.8 %						
Adv	25	3.0 %	1	2.8 %	3.8 %			
Adverbial particles	9	1.1 %						
Adv (quantifiers)	2	0.2 %						
Adv (*wh*-words)	1	0.1 %						
Adv of frequency	1	0.1 %						
Coordinating conjunctions	87	10.6 %	2	5.6 %	2.2 %			
Subordinating conjunctions	12	1.5 %						
Copulas	33	4.0 %	2	5.6 %	5.7 %	3	42.9 %	9.1 %
Definite articles	50	6.1 %	12	33.3 %	19.4 %			
Indefinite articles	12	1.5 %	2	5.6 %	14.3 %	1	14.3 %	8.3 %
Demonstrative determiners	5	0.6 %						
Determiners (numerals)	16	1.9 %						
Numerals	-							
Possessive determiners	8	1.0 %						
Determiners (quantifiers)	5	0.6 %	2	5.6 %	28.6 %			
Determiners (*wh*-words)	-							
Other determiners	-							
Existential *there*	1	0.1 %						

23

U1								
Infinitive markers	5	0.6 %						
Inserts (discourse marker, farewell, interjections, response, hesitators)	112	13.6 %						
Primary auxiliaries	22	2.7 %						
Modal auxiliaries	-							
Negators	21	2.6 %						
Common nouns	116	14.1 %	4	11.1 %	3.3 %			
Proper nouns	7	0.9 %	1	2.8 %	12.5 %			
Demonstrative pronouns	10	1.2 %						
Indefinite pronouns	9	1.1 %						
Personal pronouns (accusative)	14	1.7 %						
Personal pronouns (nominative)	63	7.7 %	4	11.1 %	6.0 %			
Interrogative pronouns	-							
Relative pronouns	5	0.6 %						
Pronouns (quantifiers)	-							
Prepositions	53	6.4 %	4	11.1 %	7.0 %			
Possessive clitics	1	0.1 %						
Lexical verbs	86	10.5 %	2	5.6 %	2.3 %	3	42.9 %	3.5 %
Total number of words	822	100.0 %	36	100.0 %	4.2 %	7	100.0 %	0.9 %

Nouns comprise 15% of the words used by this subject and therefore it is the largest category. "Inserts" is the second and "pronouns" is the third largest category in U1's speech. "Conjunctions" is the fourth and "verbs" is the fifth most frequent category here.

Existentials as well as possessive clitics are the smallest categories here. "Infinitive markers" is the second and "adverbial particles" is the third least frequent category in the material.

"Negators" is the fourth least frequent category here and "primary auxiliaries" is the fifth smallest category in this material.

The definite article category had the most omitted words in U1's material which makes up 33.3% of all the missing words here.

Subject 2. U2

Table 4. Distribution of word classes

Word classes	Provided		Omission			Incorrect		
	N	RP %	N	RP %	% of total words (provided+missing) in word class	N	RP %	% of provided in word class
Adjectives	60	5.9 %						
Adv	130	12.8 %	1	1.0 %	0.8 %			
Adverbial particles	13	1.3 %						
Adv (quantifiers)	11	1.1 %						
Adv (*wh*-words)	7	0.7 %						
Adv of frequency	-		1	1.0 %				
Coordinating conjunctions	97	9.6 %	3	3.0 %	3.0 %			
Subordinating conjunctions	3	0.3 %						
Copulas	53	5.2 %	23	23.2 %	30.3 %	1	33.3 %	1.9 %
Definite articles	33	3.3 %	12	12.1 %	26.7 %			
Indefinite articles	15	1.5 %	8	8.1 %	34.8 %	1	33.3 %	6.7 %
Demonstrative determiners	7	0.7 %						
Determiners (numerals)	32	3.2 %						
Numerals	10	1.0 %						
Possessive determiners	3	0.3 %	4	4.0 %	57.1 %			
Determiners (quantifiers)	19	1.9 %	1	1.0 %	5.0 %			
Determiners (*wh*-words)	-							
Other determiner	1	0.1 %						
Existential *there*	3	0.3 %	1	1.0 %	25.0 %			
Infinitive markers	12	1.2 %						

U2								
Inserts (discourse marker, expletives, farewell, hesitators, interjections, polite speech acts, response)	112	11.0 %	1	1.0 %	0.9 %			
Primary auxiliaries	18	1.8 %						
Modal auxiliaries	2	0.2 %				1	33.3 %	50.0 %
Negators	29	2.9 %						
Common nouns	110	10.8 %	4	4.0 %	3.5 %			
Proper nouns	12	1.2 %						
Demonstrative pronouns	35	3.5 %						
Indefinite pronouns	16	1.6 %						
Personal pronouns (accusative)	18	1.8 %	1	1.0 %	5.3 %			
Personal pronouns (nominative)	59	5.8 %	27	27.3 %	31.4 %			
Interrogative pronouns	-							
Relative pronouns	-							
Pronoun (quantifiers)	-							
Prepositions	16	1.6 %						
Possessive clitic	-							
Lexical verbs	78	7.7 %	12	12.1 %	13.3 %			
Total number of words	1014	100.0 %	99	100.0 %	8.9 %	3	100.0 %	0.3 %

"Adverbs" is the most frequent category in the material. "Pronouns" is the second and "nouns" is the third most frequent category in this subject's material. "Inserts" makes 11% of the total words and is the fourth largest category in this corpus. Conjunctions form 9.9% of the words used and make this category the fifth largest here.

The subject produced only two examples of modal auxiliaries in the corpus making this category the least frequent here. "Existentials" is the second and "Numerals" is the third least

frequent in this corpus. Infinitive markers are the fourth least frequent category in the material and adverbial particles are the fifth smallest category here.

27.3% of the omitted words in this corpus were nominative personal pronouns.

Subject 3. U3

Table 5. Distribution of word classes

Word classes	Provided		Missing			Wrong		
U3	N	RP %	N	RP %	% of total words (provided+missing) in word class	N	RP %	% of provided in word class
Adjectives	19	1.2 %						
Adv	185	11.3 %	2	1.2 %	1.1 %			
Adverbial particles	29	1.8 %	1	0.6 %	3.3 %			
Adv (quantifiers)	1	0.1 %						
Adv (*wh*-words)	1	0.1 %						
Adv of frequency	-							
Coordinating conjunctions	73	4.5 %	1	0.6 %	1.4 %			
Subordinating conjunctions	2	0.1 %						
Copulas	9	0.6 %	42	25.0 %	82.4 %			
Definite articles	92	5.6 %	58	34.5 %	38.7 %	1	9.1 %	1.1 %
Indefinite articles	8	0.5 %	12	7.1 %	60.0 %			
Demonstrative determiners	9	0.6 %	2	1.2 %	18.2 %			
Determiners (numerals)	31	1.9 %						
Numerals	81	5.0 %						
Possessive determiners	1	0.1 %						
Determiners (quantifiers)	8	0.5 %	1	0.6 %	11.1 %			
Determiners (*wh*-words)	1	0.1 %						
Other determiner	-							
Existential *there*	1	0.1 %	1	0.6 %	50.0 %			
Infinitive markers	1	0.1 %	1	0.6 %	50.0 %			

Inserts (discourse marker, farewell, interjections, response, hesitators)	230	14.1 %						
Primary auxiliaries	39	2.4 %	1	0.6 %	2.5 %			
Modal auxiliaries	-		1	0.6 %				
Negators	62	3.8 %				1	9.1 %	1.6 %
Common nouns	278	17.1 %	1	0.6 %	0.4 %			
Proper nouns	12	0.7 %						
Demonstrative pronouns	-							
Indefinite pronouns	19	1.2 %						
Personal pronouns (accusative)	5	0.3 %	1	0.6 %	16.7 %	2	18.2 %	40.0 %
Personal pronouns (nominative)	153	9.4 %	20	11.9 %	11.6 %			
Interrogative pronouns	3	0.2 %						
Relative pronouns	-							
Pronouns (quantifiers)	1	0.1 %						
Prepositions	47	2.9 %	16	9.5 %	25.4 %	1	9.1 %	2.1 %
Possessive clitics	2	0.1 %						
Lexical verbs	227	13.9 %	7	4.2 %	3.0 %	6	54.5 %	2.6 %
Total number of words	1630	100.0 %	168	100.0 %	9.3 %	11	100.0 %	0.7 %

"Nouns" constitutes 17.8% of the total number of words and is the largest category in this corpus. Inserts make up 14.1% of the total number of words and as a result are the second largest category in the material. Lexical verbs are the third most frequent in the corpus. All the adverbs together constitute 11.5% of words used and are the fourth largest category in this subject's material. All the pronouns form 11.2% of the total words and make the fifth largest category in this corpus.

Adverbial particles are the fifth least frequent in this corpus. "Adjectives" is the fourth smallest category in U3's material. Copulas are the third least frequent category here. "Possessive clitics" contains only 0.1% of all the words and is the second least frequent

category here. Existentials, along with infinitive markers, are the least frequent categories here.

Subject 4. H1

Table 6. Distribution of word classes

Word classes	Provided		Missing			Wrong		
H1								
	N	RP %	N	RP %	% of total words (provided+missing) in word class	N	RP %	% of provided in word class
Adjectives	15	3.2 %						
Adv	10	2.2 %						
Adverbial particles	4	0.9 %						
Adv (quantifiers)	1	0.2 %						
Adv (*wh*-words)	4	0.9 %						
Adv of frequency	-							
Coordinating conjunctions	24	5.2 %						
Subordinating conjunctions	1	0.2 %						
Copulas	18	3.9 %	8	19.5 %	30.8 %	1	50.0 %	5.6 %
Definite articles	4	0.9 %	8	19.5 %	66.7 %			
Indefinite articles	4	0.9 %	1	2.4 %	20.0 %			
Demonstrative determiners	1	0.2 %						
Determiners (numerals)	-							
Numerals	22	4.7 %						
Possessive determiners	2	0.4 %	1	2.4 %	33.3 %			
Determiners (quantifiers)	1	0.2 %						
Determiners (*wh*-words)	-							
Other determiners	-							
Existential *there*	-							
Infinitive markers	6	1.3 %						

H1								
Inserts (discourse marker, expletives, greetings, hesitators, interjections, polite speech acts, responses)	131	28.2 %						
Primary auxiliaries	15	3.2 %						
Modal auxiliaries	18	3.9 %						
Negators	24	5.2 %						
Common nouns	22	4.7 %	6	14.6 %	21.4 %			
Proper nouns	4	0.9 %						
Demonstrative pronouns	6	1.3 %						
Indefinite pronouns	-							
Personal pronouns (accusative)	4	0.9 %	1	2.4 %	20.0 %			
Personal pronouns (nominative)	49	10.5 %	15	36.6 %	23.4 %			
Interrogative pronouns	6	1.3 %						
Relative pronouns	-							
Pronouns (quantifiers)	-							
Prepositions	2	0.4 %						
Possessive clitics	-							
Lexical verbs	67	14.4 %	1	2.4 %	1.5 %	1	50.0 %	1.5 %
Total number of words	465	100.0 %	41	100.0 %	8.1 %	2	100.0 %	0.4 %

Inserts are the largest category in this corpus. "Lexical verbs" is the second and "pronouns" is the third most frequent category here. "Nouns" is the fourth largest category in H1's material and "conjunctions" is the top five word class in this subject's material.

Prepositions are the least frequent category here. Adverbial particles and determiners are among the second smallest categories in the material. Infinitive markers are the third least frequent category in the corpus. "Articles" is the bottom four category and "adverbs" is the fifth least frequent category in this subject's material.

Subject 5. H2

Table 7. Distribution of word classes

Word classes	Provided		Missing			Wrong		
	N	RP %	N	RP %	% of total words (provided+missing) in word class	N	RP %	% of provided in word class
Adjectives	3	0.6 %						
Adv	4	0.8 %						
Adverbial particles	3	0.6 %						
Adv (quantifiers)	-							
Adv (*wh*-words)	-							
Adv of frequency	-							
Coordinating conjunctions	12	2.4 %						
Subordinating conjunctions	-							
Copulas	13	2.6 %	1	3.7 %	7.1 %			
Definite articles	85	16.9 %	9	33.3 %	9.6 %			
Indefinite articles	6	1.2 %	2	7.4 %	25.0 %	1	50.0 %	16.7 %
Demonstrative determiners	-							
Determiners (numerals)	8	1.6 %						
Numerals	-							
Possessive determiners	3	0.6 %						
Determiners (quantifiers)	-							
Determiners (*wh*-words)	-							
Other determiners	-							
Existential *there*	1	0.2 %						
Infinitive markers	-		1	3.7 %				
Inserts (discourse marker, farewell, interjections, response, hesitators)	88	17.5 %						

H2								
Primary auxiliaries	36	7.1 %						
Modal auxiliaries	-							
Negators	16	3.2 %						
Common nouns	101	20.0 %	7	25.9 %	6.5 %			
Proper nouns	1	0.2 %						
Demonstrative pronouns	-							
Indefinite pronouns	-							
Personal pronouns (accusative)	4	0.8 %						
Personal pronouns (nominative)	27	5.4 %	1	3.7 %	3.6 %			
Interrogative pronouns	2	0.4 %						
Relative pronouns	-							
Pronouns (quantifiers)	-							
Prepositions	24	4.8 %	3	11.1 %	11.1 %			
Possessive clitics	-							
Lexical verbs	67	13.3 %	3	11.1 %	4.3 %	1	50.0 %	1.5 %
Total number of words	504	100.0 %	27	100.0 %	5.1 %	2	100,0 %	0.4 %

Nouns are the largest category in this corpus. "Articles" is the second and "Inserts" is the third most frequent category in this corpus. "Lexical verbs" constitutes 13.3% of the total words and is top four categories in this material. "Primary auxiliaries" is the fifth largest category in H2's corpus.

Existentials are the smallest category in this material. Adjectives as well as adverbial particles are the second least frequent categories here. "Adverbs" is the third and "determiners" is the fourth least frequent category in H2's corpus. "Conjunctions" is the fifth least frequent word class here.

Subject 6. H3

Table 8. Distribution of word classes

H3								
Word classes	Provided		Missing			Wrong		
	N	RP %	N	RP %	% of total words (provided+missing) in word class	N	RP %	% of provided in word class
Adjectives	14	3.0 %						
Adv	8	1.7 %	1	2.8 %	11.1 %			
Adverbial particles	5	1.1 %	1	2.8 %	16.7 %			
Adv (quantifiers)	-							
Adv (*wh*-words)	2	0.4 %						
Adv of frequency	-							
Coordinating conjunctions	32	6.8 %						
Subordinating conjunctions	4	0.9 %						
Copulas	13	2.8 %	7	19.4 %	35.0 %			
Definite articles	45	9.6 %	5	13.9 %	10.0 %			
Indefinite articles	7	1.5 %	1	2.8 %	12.5 %	1	50.0 %	14.3 %
Demonstrative determiners	1	0.2 %						
Determiners (numerals)	7	1.5 %						
Numerals	7	1.5 %						
Possessive determiners	3	0.6 %						
Determiners (quantifiers)	-							
Determiners (*wh*-words)	-							
Other determiner	-							
Existential *there*	-							
Infinitive markers	11	2.4 %						

H3								
Inserts (hesitators, interjections, polite speech acts, response)	46	9.8 %						
Primary auxiliaries	21	4.5 %						
Modal auxiliaries	1	0.2 %						
Negators	4	0.9 %						
Common nouns	65	13.9 %	1	2.8 %	1.5 %			
Proper nouns	10	2.1 %						
Demonstrative pronouns	-							
Indefinite pronouns	-							
Personal pronouns (accusative)	7	1.5 %	1	2.8 %	12.5 %			
Personal pronouns (nominative)	62	13.2 %	12	33.3 %	16.2 %			
Interrogative pronouns	1	0.2 %						
Relative pronouns	1	0.2 %						
Pronouns (quantifiers)	4	0.9 %						
Prepositions	10	2.1 %	1	2.8 %	9.1 %			
Possessive clitics	-							
Lexical verbs	77	16.5 %	6	16.7 %	7.2 %	1	50.0%	1.2 %
Total number of words	468	100.0 %	36	100.0 %	7.1 %	2	100.0 %	0.4 %

"Lexical verbs" is the largest category in this corpus. "Nouns" constitutes 16% of all the words and is the second most frequent category along with pronouns here. "Articles" forms 11.1 % of the total number of words and is the third most frequent category. "Inserts" is the fourth and "conjunctions" is the fifth most frequent word class in this subject's material.

"Modal auxiliary" is the least frequent category in the corpus. Negators are the second and Adverbial particles are the third least frequent word classes in the corpus. "Numerals" is the fourth and "adverbs" is the fifth smallest category in this corpus.

3.1.1.2 Distribution of Grammatical Categories Within Participants of This Study

Table 9. The five most and least frequent grammatical categories are highlighted for each participant

Categories	University graduates									High school graduates								
	U1			U2			U3			H1			H2			H3		
	R	N	%	R	N	%	R	N	%	R	N	%	R	N	%	R	N	%
Nouns	1	123	15.0	3	122	12.0	1	290	17.8	4	26	5.6	1	102	20.2	2	75	16.0
Inserts	2	112	13.6	4	112	11.0	2	230	14.1	1	131	28.2	3	88	17.5	4	46	9.8
Lexical verbs	5	86	10.5	6	78	7.7	3	227	13.9	2	67	14.4	4	67	13.3	1	77	16.5
Adverbs	11	29	3.4	1	148	14.6	4	187	11.5	9	15	3.3	12	4	0.8	9	10	2.1
Pronouns	3	101	12.3	2	128	12.7	5	181	11.2	3	65	14.1	6	33	6.6	2	75	16.0
Articles	6	62	7.6	10	48	4.8	6	100	6.1	10	8	1.8	2	91	18.1	3	52	11.1
Numerals	-	-	-	16	10	1.0	7	81	5.0	7	22	4.7	-	-	-	10	7	1.5
Conjunctions	4	99	12.1	5	100	9.9	8	75	4.6	5	25	5.4	10	12	2.4	5	36	7.7
Copulas	9	33	4.0	9	53	5.2	15	9	0.6	8	18	3.9	9	13	2.6	7	13	2.8
Negators	13	21	2.6	11	29	2.9	9	62	3.8	6	24	5.2	8	16	3.2	12	4	0.9
Adverbial particles	14	9	1.1	14	13	1.3	13	29	1.8	12	4	0.9	13	3	0.6	11	5	1.1
Adjectives	10	31	3.8	8	60	5.9	14	19	1.2	9	15	3.2	13	3	0.6	7	14	3.0
Primary auxiliaries	12	22	2.7	12	18	1.8	12	39	2.4	9	15	3.2	5	36	7.1	6	21	4.5
Determiners	8	34	4.1	7	62	6.2	10	50	3.2	12	4	0.8	11	11	2.2	8	11	2.4
Infinitive markers	15	5	0.6	15	12	1.2	17	1	0.1	11	6	1.3	-	-	-	8	11	2.4
Possessive clitics	16	1	0.1	-	-	-	16	2	0.12	-	-	-	-	-	-	-	-	-
Prepositions	7	53	6.4	13	16	1.6	11	47	2.9	13	2	0.4	7	24	4.8	9	10	2.1
Modal auxiliaries	-	-	-	18	2	0.2	-	-	-	8	18	3.9	-	-	-	13	1	0.2

Existentials	16	1	0.1	17	3	0.3	17	1	0.1	-	-	-	14	1	0.2	-	-	-

The table above presents the distribution of word classes along with the total number of words in that category, the relative distribution of that category as a percentage of the total words and the rank of that category in each subject's corpus. The table shows the five most frequent categories with grey and the five least frequent categories with black for each subject.

3.1.1.2.1 University graduates

3.1.1.2.1.1 The five most frequent categories among university graduates

As shown in table 9, U1 and U3 share the same two most frequent categories which are nouns and inserts respectively. The third most frequent category in U1's corpus is pronouns while the third largest category in U3's material is lexical verbs.

Unexpectedly, "adverbs" is the largest category in U2's corpus while it is the fourth top category for U3.

U2 has used pronouns more than nouns. Therefore, "pronouns" is the second most frequent category and "nouns" is the third largest category in U2's corpus.

"Inserts" was the second largest category for U1and U3 while it is the fourth largest category for U2.

The fourth category in U1's material is equivalent to U2's fifth top category which is conjunctions. "Lexical verbs" is U1's fifth most frequent category and "pronouns" is U3's fifth largest category.

3.1.1.2.1.2 The five least frequent categories among university graduates

According to table 9, U3's and U1's two least frequent categories are the same, i.e. possessive clitics, existentials and infinitive markers. However, they are not in the same order. Both U1 and U3 produced only one existential each. "Existentials" is in the bottom two category for U2 with three examples. "Existentials" is the only category which is among one of the three least frequent categories for all university graduates here.

One possessive clitic was used by U1and two instances of this category were produced by U3. No example of this category was found in U2's corpus.

Another least frequent category in U3's material is infinitive marker with one example. Five cases of "infinitive markers" make it the second smallest category in U1's speech. "Infinitive markers" is the fourth least frequent category in U2's material with twelve instances.

U2's smallest category is "modal auxiliaries" with only two cases. This category was not found in any of the other university graduate subjects.

The third least frequent category in U1 is "adverbial particles". This is the fifth least frequent category both for U3 and U2.

U2's third smallest category is numerals. U1 has not used this category. She has used numerals only as determiners.

U3's third smallest category is copulas. The other two university graduates have used this category more often. "Adjectives" is also the fourth least common category in U3's corpus while this category is more frequent for the other two subjects.

"Negators" is the fourth least frequent category in U1's corpus. This category is more common in the other two university graduate subjects' corpora which mean that they have used more negative sentences than U1.

Comparing university graduates with one another U1's and U3's top two most frequent categories are exactly the same and their bottom two categories are also the same but not in the same order.

Without considering the rankings, nouns, lexical verbs, inserts and pronouns are among the five top categories for U1 and U3. All these mentioned categories except lexical verbs are also among one of the four most frequent categories in U2's corpus.

3.1.1.2.2 High school graduates

3.1.1.2.2.1 The five most frequent categories among high school graduates

The most frequent category in H1's corpus is inserts with 28.2% of total number of words. "Nouns" is the largest category in H2's material. "Lexical verbs" with 16.5% of all the words is the largest category produced by H3. "Nouns" with 16% of total words is the second largest category in H3's corpus, which means that the difference between the first and the second largest category for this subject differs by only two numbers. Lexical verbs are the second most frequent category for H1. The difference between the first and second most frequent

categories for this participant is almost half of the total number of words, which is a distinct difference. After "nouns", "articles" is the largest category used by H2. The third largest category found in H1's material is pronouns while inserts for H2 and articles for H3 have the same rankings.

"Nouns" is the fourth most frequent category in H1's corpus. Lexical verbs and inserts are the fourth largest categories produced by H2 and H3 respectively.

"Conjunctions" is the fifth largest category in both H1's and H3's corpus while "primary auxiliaries" has the same ranking in H2's material. "Conjunctions" is also H2's fifth least frequent word class which is a big difference comparing to H1's and H3's corpora.

Nouns, lexical verbs and inserts are among one of the four most frequent categories for all high school graduates in this study.

A significant difference between these three subjects is the usage of articles. It is the second and third most frequent category in H2 and H3's material consequently while H1 used it as his fourth least frequent category.

3.1.1.2.2.2 *The five least frequent categories among high school graduates*
"Prepositions" is the smallest category in H1's speech while existentials and modal auxiliaries with only one example each are the least frequent categories in H2's and H3's corpora respectively. "Determiners" is among the second least frequent category found in H1's and the fourth smallest category in H2's material. The second least frequent category in H3's corpus is negators whereas adverbial particles and adjectives are the second smallest word classes in H2's material. The third least frequent category produced by H1 is infinitive markers while adverbs for H2 and adverbial particles for H3 have the same ranking. "Articles" is the fourth least frequent category used by H1 whereas determiners and numerals have the same ranking for H2 and H3 consecutively. The fifth smallest category found in H1's corpus are adjectives and primary auxiliaries whereas conjunctions in H2's material and adverbs in H3's corpus have the same place.

3.1.1.3 The Distribution of three main word classes

Table 10. Three main word classes used by all the subjects

Word classes	University graduates						High school graduates					
	U1		U2		U3		H1		H2		H3	
	N	%	N	%	N	%	N	%	N	%	N	%
Functional words	444	54.0	512	50.5	679	42.0	216	46.5	256	49.0	247	53.0
Lexical (content) words	266	32.0	390	38.5	721	44.0	118	25.5	176	34.0	175	37.0
Inserts	112	14.0	112	11.0	230	14.0	131	28.0	88	17.0	46	10.0

U1 used 54% functional words, 32% lexical words and 14% inserts in her speech. U2 produced 50.5 % functional words, 38.5 % lexical words and 11% inserts. 49% functional words, 34% content words and 17% inserts were found in H2's corpus. 53% functional words, 37% lexical words followed by 10% inserts were in H3's material. The subjects discussed so far have used between 49% to 54% functional words in their speech. Therefore, "functional words" is the largest word class in these four subjects' corpora. "Lexical words" with minimum 32% and maximum 38.5%, is the second largest word class and "inserts" between 10 and 17 per cent, is the least frequent word class for these four participants.

One of the subjects, U3, from the university graduates has used 44% lexical words, 42% functional words and 14% inserts. As a result "content words" is the largest category and "functional words" is the second most frequent category in his corpus. For this subject, "inserts" is also among the least frequent category the same as the other four subjects mentioned earlier.

H1 produced 46.5% functional words. So this category is the most frequent category for U1, U2, H2 and H3. However, his second largest category unlike any other subjects in this study is inserts with 28.2%. And as a result his least frequent category is content words with 25.5%. This means he has used almost double number of functional words than content words.

Table 11. Total number of words and mean of the sentences

	University graduates				High school graduates			
	U1	U2	U3	University graduates	H1	H2	H3	High school graduates
Total number of words	822	1014	1630	3466	465	520	468	1437
Mean of sentences	6.94	4.61	5.70	5.55	2.54	5.61	4.57	3.88

3.1.1.3.1 Word frequencies and sentence length

Table 11 demonstrates the total number of words and mean of the sentences produced by each participant and each group. The subject who used the highest number of words is U3 with 1630. The second subject is U2 with 1014 words. The third person is U1 with 822 words. These three subjects are all among the university graduate group. In the other group, H2 used 520 words as the fourth subject using the most number of words. The fifth subject in this study who is the second subject among high school graduates is H3 with 468 words and the last subject who used the least amount of words both among high school graduates and among all the subjects in this investigation is H1 with 465 words. However, it must be noted that the difference between H3's and H1's word productions is only three words.

The mean length of the sentences produced by each subject is compared not only within the subgroups but also compared between the two groups. The participant with the longest sentence length in this study is U1 with a mean of 6.94. U3 is the second subject with more words in his sentences in this paper. He has a mean of 5.70. H2 is the third participant of this research with longest mean of the sentences (5.61) in his sentences. U2 has the fourth longest mean of sentences (4.61) among these subjects. H3 is the fifth subject with longest mean of sentences in her sentences (4.57) and H1 with the mean of 2.54 has the shortest mean of sentences in his sentences in this investigation.

3.1.2 Morphological Analysis

3.1.2.1 The Distribution of inflectional Morphemes

Table 12. Distribution of inflectional morphemes within the subjects

Inflectional Morphemes	University graduates				High school graduates			
	U1 N	U2 N	U3 N	All university graduates together	H1 N	H2 N	H3 N	All high school graduates together
Third person singular present	23	1	15	39	-	1	-	1
Regular verb past tense marker	4	2	34	40	-	1	5	6
Irregular verb past tense marker	13	1	19	33	7	-	15	22
Ing marker	14	10	36	60	11	28	20	59
Possessive marker	1	-	2	3	-	-	-	-
Regular plural nouns	11	31	21	63	4	10	14	28
Irregular plural nouns	3	7	4	14	1	-	-	1
Total Number of Morphemes	69	52	131	252	23	40	54	117
Relative distribution to total number of words	8.3%	5.1%	8.0%	7.27	4.9%	7.6%	11.5%	8.14

In this section, all the different inflectional morphemes used by each subject are calculated and presented in table 12.

3.1.2.1.1 University graduates

Subject 1. U1

As shown in table 12, U1 has used verb present morpheme *s* as in *he works* 23 times which is the most frequent morpheme for this subject. The second most frequent inflectional morpheme is the *ing* marker with 14 examples. "Irregular verb past tense markers" as *found* in *she found* is the third largest morpheme here. "Regular plural noun markers" is the fourth most frequent morpheme with 11 examples. Regular verb past tense markers are the third least frequent category here. "Irregular plural nouns" with three cases is the second from bottom. The least common inflectional morpheme used by U1 is possessive marker with only one example. 8.3% of the total number of words used by this participant have inflectional morphemes.

Subject 2. U2

This subject used 31 "regular plural nouns" which makes it the most frequent morpheme in his corpus. Ten *ing* markers and seven irregular plural noun morphemes were also found in U2's material. Two regular verb past tense markers, one irregular verb past tense marker and one verb present morpheme were also produced. The subject did not use any possessive marker in his speech.

Subject 3. U3

Thirty six *ing* markers, 34 regular verb past tense markers, 21 regular plural nouns and 19 irregular verb past tense markers were used by U3. He also produced 15 verb present morphemes, four irregular plural nouns and two possessive markers.

3.1.2.1.2 High school graduates

Subject 4. H1

Eleven *ing* markers, seven irregular verb past tense markers, four regular plural noun morphemes and only one irregular plural noun morpheme were used. Verb present morpheme, regular verb past tense marker and possessive markers were not found in this subject's corpus.

Subject 5. H2

There were 28 *ing* markers, ten regular plural noun morphemes, one verb present morpheme and one regular verb past tense marker in H2's corpus. On the other hand, irregular verb past tense marker, possessive marker and irregular plural noun morphemes were not found in this material.

Subject 6. H3

This subject produced five regular verb past tense markers, 15 irregular verb past tense markers, 20 *ing* markers and 14 regular plural nouns. Verb present morpheme, passive marker as well as irregular plural noun morphemes were not used by H3.

3.1.2.2 Morphological errors

Morphological errors made by each participant will be discussed separately here. The explanations of error codes are presented in the method section.

3.1.2.2.1 University graduates

Subject 1. U1

[*m:0es]

The verb present morpheme *es* is missing after *go* in the following sentence *go to the ball but this person doesn't go, I don't think*. This is the only morphological error this participant had in her speech.

Subject 2. U2

[*m:0ed]

One past tense marker *ed* was required instead of *ing* marker.

[*m:0es]

In two cases the verb present morpheme *es* were missing after *go*.

One *s* verb present marker is left out after *mist* in this example *oh my goodness it mist*.

[*m:0ing]

The *ing* marker was missing after *mow* in the following sentence *mow the grass*.

[*m:a]

In two other instances, the subject uses the singular form of the verb *is* after the plural noun *dresses*. These are also morphological errors as the subject and verb do not agree.

Subject 3. U3

[*m:0ed]

Three *ed* regular verb past tense markers were missing in the material.

[*m:0s]

One regular plural noun maker *s* was required but not provided in the corpus.

[*m:0es]

Three verb present morpheme *s* were missing in this corpus.

[*m:+ing]

One *ing* marker was provided in an example where it was not required.

3.1.2.2.2 High school graduates

Subject 4. H1

[*m:0ed]

One regular verb past tense marker *ed* was missing in this subject's corpus.

Subject 5. H2

[*m:0es]

One verb present morpheme *s* was missing.

[m:a:+s]

The regular plural noun morpheme *s* was not required in the provided example.

Subject 6. H3

[*m:0ed]

In five examples, regular verb past tense *ed* markers were missing.

[*m:0es]

Four instances verb present morpheme *s* were not provided.

[m:a:+s]

Regular plural noun marker *s* was not needed in one of the supplied cases.

3.1.3 Semantic Analysis

In this section, semantic error made by each subject during his/her interview will be analyzed separately. The explanations of error codes are presented in the method section.

Subject 1. U1

[* s:ur]

Copular verb *was* was produced instead of lexical verb *use* in one case.

[* s:r]

This subject had nine examples of this type of semantic error in her speech. In four examples, another lexical verb was required instead of the lexical verb provided by the subject. In five instances, the wrong preposition was supplied.

Subject 2. U2

[* s:r]

In one instance, adverb *ago* was supplied instead of *later*. In three examples, different common nouns were needed than the ones provided.

[* s:per]

This subject used three examples of this kind of semantic error. Three common nouns were used in sentences that did not make much sense. For example *the water is water [* s:per]*. The second *water* makes this sentence completely meaningless.

[* s:uk]

U2 used one common noun *ham* as well as three lexical verbs *flew*, *driving* and *listening* where these words did not make any sense given the context in which they were used.

Subject 3. U3

[* s:ur]

Nominative personal pronoun *I* was used twice instead of the preposition *in*. In the third example, it is the opposite. The preposition *in* was needed instead of personal pronoun *I*. Copular verb *is* should have been used instead of preposition *in*.

Indefinite article *a* was produced incorrectly where the copular verb *is* should have been supplied.

In another case, coordinating conjunction *but* was used instead of the preposition *with*.

This type of semantic error was found in three common nouns. One lexical verb *look* instead of copular verb *was* was found in this corpus.

[* s:r]

This kind of semantic error occurred in four common nouns, one lexical verb and one preposition supplied by the participant. For example an instance of *castle mother* for *queen* was found in the common nouns.

[* s:per]

This type of semantic error was found in one common noun.

Subject 4. H1

[* s:ur]

In one example indefinite article *a* was used instead of nominative personal pronoun *I*.

Modal auxiliary *can* was provided instead of the modal *do* in another instance.

[* s:r]

In one case, adverb *next* was produced for *last*.

[* s:per]

Two common nouns were found to have this kind of semantic errors.

[* s:uk]

Two examples of common nouns supplied involved this type of semantic error.

Subject 5. H2

[* s:ur]

In two of the following sentences *the ball is in the window for um the dude* and *the dude is in the window* lexical verb *hit* should have been used instead of the copular verb *is* and the preposition *in*.

In addition, some prepositions are used incorrectly. In one case, preposition *in* is used instead of lexical verb *hit*. In another instance, preposition *by* is used instead of *for*.

[* s:r]

Four common nouns, one nominative personal pronoun and two lexical verbs were involved in this kind of semantic errors.

[* s:per]

Twenty common nouns were found with this type of semantic error in the corpus.

[* s:uk]

One lexical verb as well as two common nouns in this material had semantic error of this kind.

Subject 6. H3

[* s:ur]

In one instance, the copular *was* was used instead of lexical verb *forgot*. In two other examples the adjective *scared* was used instead of *sad*. The common noun *window* for *father* and definite article and common noun *the boy* instead of personal pronoun *it* were used incorrectly.

[* s:r]

One adjective, three lexical verbs, four common nouns, 20 nominative personal pronouns, one accusative personal pronoun and two possessive determiners were involved in semantic error here.

[* s:per]

Lexical verb *go* was a semantic error in one instance.

[* s:uk]

Two lexical verbs, one adjective, one common noun and two proper nouns were among the semantic errors.

3.1.4 Phonological Analysis

Phonological errors made by the participants of this study will be investigated here. The explanations of error codes are presented in the method section.

Subject 1. U1

[* p:w]

Two examples of real word phonological errors *supposedly* and *and* occurred in this corpus.

[* p:n]

Two instances of non-word phonological errors *gunning* and *глok@u* were found in this material.

Subject 2. U2

No phonological errors were found in this subject's material.

Subject 3. U3

[* p:w]

Four real word phonological errors were provided by this subject.

[* p:n]

Eleven of the phonological errors supplied by this subject were non-words.

[* p:m]

One phonological error *faɪɹə@u* of this kind was used by U3.

Subject 4. H1

[* p:n]

Two non-word phonological errors *mat* and *jɛts@u* were provided by H1.

Subject 5. H2

[* p:w]

In one case the subject used *mate* which is counted as a real word phonological error.

[* p:n]

In three instances, the participant used the non-word *brover* which is a non-word phonological error.

Subject 6. H3

[* p:w]

Thirteen examples of real word phonological errors were found in this corpus.

[* p:n]

Six cases of non-word phonological errors were used by H3.

3.1.5 Lexical Analysis

3.1.5.1 WordSmith Findings

WordSmith tools has been used to investigate the mean length of the sentences and to compute type-token ratio (TTR). "Mean of the sentences" is discussed in the syntactic analysis but TTR will be presented here.

Table 13. TTR values for each participant.

	University graduates			High school graduates		
	U1	U2	U3	H1	H2	H3
TTR	28.92	18.06	19.33	27.25	21.69	33.76

3.1.5.1.1 Type Token Ratio

H3 had the highest TTR (33.76) among the subjects. U1 is the second participant with the highest TTR (28.92) here. The third participant with the highest TTR (27.25) is H1. H2 is the fourth subject with TTR of 21.69. U3 is in the bottom two with the TTR of 19.33 and U2 had the lowest TTR (18.06) in this study.

According to *Longman Grammar of Spoken and Written English* (Biber et al. 1999, pp.52-53) the relationship between the number of different word forms (distinct words) or type and the number of running words (token) is called the type-token ratio (TTR). The more types (distinct words) there are compared to the number of tokens, the more varied the vocabulary is, i.e. there is greater lexical variety (Williamson, 2009).

3.1.5.2 Neologistic errors

The neologistic errors found in the subjects' corpora will be studied here. The explanations of error codes are presented in the method section.

Subject 1. U1

[* n:k]

One example of non-word neologistic error *ɪmbɪləteɪɪd@u* was found in U1's speech.

Subject 2. U2

[* n:k]

There is one instance *hæmbɹə@u* in the subject's corpus.

[* n:uk]

Two cases of this type of neologistic error *dədɪl@u* and *pɹɛʃɪn@u* were produced in this material.

Subject 3. U3

[* n:k]

There were 14 examples of this sort of neologistic error in the corpus.

[* n:uk]

One instance of neologistic error *kamənaz@u* in which the target was unknown was found.

Subject 4. H1

[* n:k]

Neologistic error *bɹɛlə@u* was used in one instance.

[* n:uk]

Two other examples of neologistic errors *sebɚ@u* and *ærə@u* were also found.

Subject 5. H2

[* n:k]

This neologistic error *pɹaɪətstɹupə* was supplied by the subject.

[* n:uk]

glæsɪt@u is another neologistic error in this material.

Subject 6. H3

There was no neologistic error in this subject's corpus.

3.2 Interpretation of the Results

In this section, a summary and discussion of the results from all the subjects in both of the groups, high school graduates and university graduates, will be compared with one another and some of the results of this study will be compared with the Longman spoken corpora.

3.2.1 Summary of Syntactic analysis

3.2.1.1 Distribution of Grammatical Categories within participants

There is a difference regarding the number of different word classes each subject supplied in this study.

Table 14. Word classes not produced by one or some of the subjects.

	University graduates			High school graduates		
Word classes	U1	U2	U3	H1	H2	H3
Adv of frequency		-	-	-	-	-
Numerals	-				-	
Modal auxiliary	-		-		-	
Interrogative pronoun	-	-				
Relative pronoun		-	-	-	-	
Indefinite pronoun				-	-	-
Demonstrative pronoun			-		-	-
Demonstrative determiner					-	
Subordinating conjunction					-	
Existential there				-		-
Possessive clitics		-		-	-	-
Infinitive marker					-	

Table 14 shows word classes which are not used by one or some of the subjects in this investigation. According to the table presented above, U1 did not produce three of the 29

main categories in this paper. U2 and U3 did not use four of those 29 word classes each. H1 and H3 used 24 out of 29 word classes each. H2 is the subject who used the least number of word classes in this study with 19 out of 29.

3.2.1.2 Omissions and incorrect usage of word classes

Table 15. Omission and incorrect usage of word classes among participants

	University graduates					High school graduates				
	U1	U2	U3	Total		H1	H2	H3	Total	
Grammatical categories	RP%	RP%	RP%	Average RP%	Std dev	RP%	RP%	RP%	Average RP%	Std dev
Omission (provided+missing)	4.2 %	8.9 %	9.3 %	7.5 %	2.3 %	8.1 %	5.1 %	7.1 %	6.8 %	1.2 %
Incorrect (of provided)	0.9 %	0.3 %	0.7 %	0.6 %	0.2 %	0.4 %	0.4 %	0.4 %	0.4 %	0.0 %

According to the above table, university graduates had an omission between 4.2 and 9.3 per cent. On the other hand, high school graduates omitted the required words between 5.1% and 8.1%. Therefore, all the participants in this study have left out less than 10% of the required words. The subject who had the least omitted words is U1, with relative proportion of 4.2%; U3 with 9.3% is the subject with the most omitted words in this paper. The average relative proportion of the missing words for university graduates is 7.5% with standard deviation of 2.3%. On the other hand, the average relative proportion of the omitted words of the high school graduate group is 6.8% with standard deviation of 1.2%. This implies that on average, first, university graduates omitted 0.7% more words than high school graduates and second, there is a bigger difference within the university graduate group than for the high school graduates in the number of omissions.

Taking incorrect usage of word classes into account, U2 who is the only person with Moderate aphasia in this investigation had the least incorrect examples in his corpus. U1 has produced the most incorrect instances here. However as the incorrect usage of word classes is between 0.3% and 0.9% it implies that all of the subjects used the words which could be investigated in this paper more than 99% of the times correctly. University graduates have used the word classes incorrectly 0.2% more than high school graduates. As a result they made one third more mistakes, on average, than high school graduates.

3.2.1.3 Similarities between university graduates and high school graduates regarding the five most and least frequent word classes

Nouns and inserts are among two of the five most frequent categories for all the subjects. All the subjects except U2 have lexical verbs among one of the largest five categories in this study. "Pronouns" is among one of the five most frequent categories for all the participants except H2. "Conjunctions" is among either fourth or fifth largest category for four of the subjects: U1, U2, H1 and H3.

"Adverbial particles" is among one of the five bottom categories for all the subjects here.

"Infinitive markers" is among either second, third or fourth least frequent category for four subjects of this study: U1, U2, U3 and H1.

"Existentials" is the first least frequent category in U1, U3 and H2's corpora. It is the second least frequent in U2's material.

"Modal auxiliaries" is the first least frequent category for U2 and H3.

3.2.1.4 Differences between university graduates and high school graduates regarding the five most and least frequent word classes

All the university graduates used at least one example of existential *there* in their speech while only one of the high school graduates, H2, has produced an instance of this category. All university graduates have used infinitive markers whereas only H1 has produced this category among high school graduates. "Possessive clitics" is found in only two corpora, U1 and U3, in this investigation. "Prepositions" is the smallest category in H1's corpus while other subjects have used this category more often and as a result it is not even among one of their fifth least frequent categories. Another difference is among the usage of adverbs. H2 has this category as the third least frequent, H3 and H1 as the fifth least frequent in their materials while it is among the first and fourth most frequent category for U2 and H3 respectively. This means that for all the high school graduates this was either the third or fifth least frequent category, however it is the first and fourth most frequent word class for two of the university graduates.

3.2.1.5 The comparison of word class distributions between Longman's corpus and the findings from this study
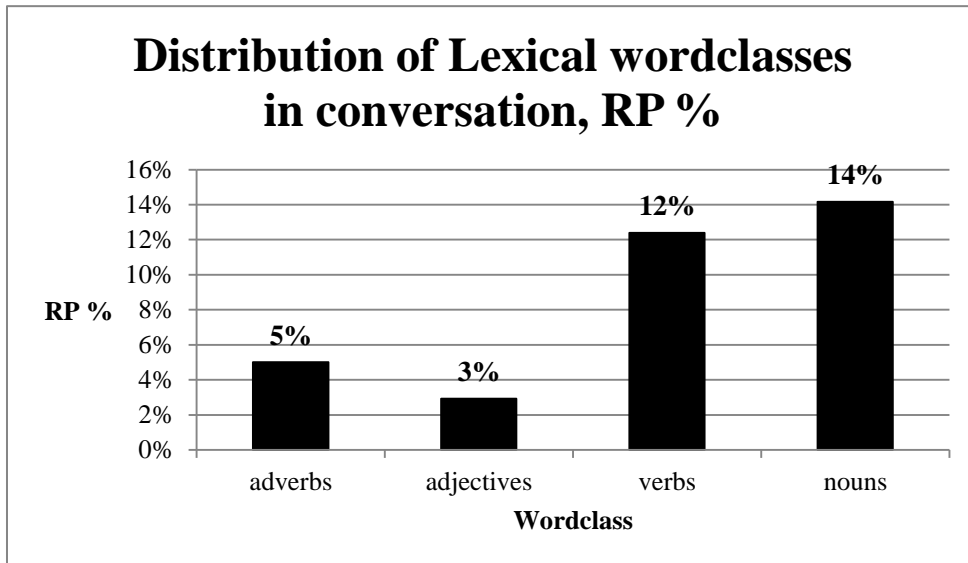
3.2.1.5.1 Distribution of word classes in the Longman spoken corpus

The distribution of all word classes in this study is not presented in Longman. For example, there is no information about the frequency of negators, existentials, possessive clitics and infinitive markers in Longman and on the other hand, these four categories are among the least frequent categories in this study. Therefore, the comparison will be made only between the five most frequent categories in Longman's corpus and this paper's results.

The following graphs demonstrate the distribution of inserts, lexical and functional word classes in conversation in the Longman spoken corpus.
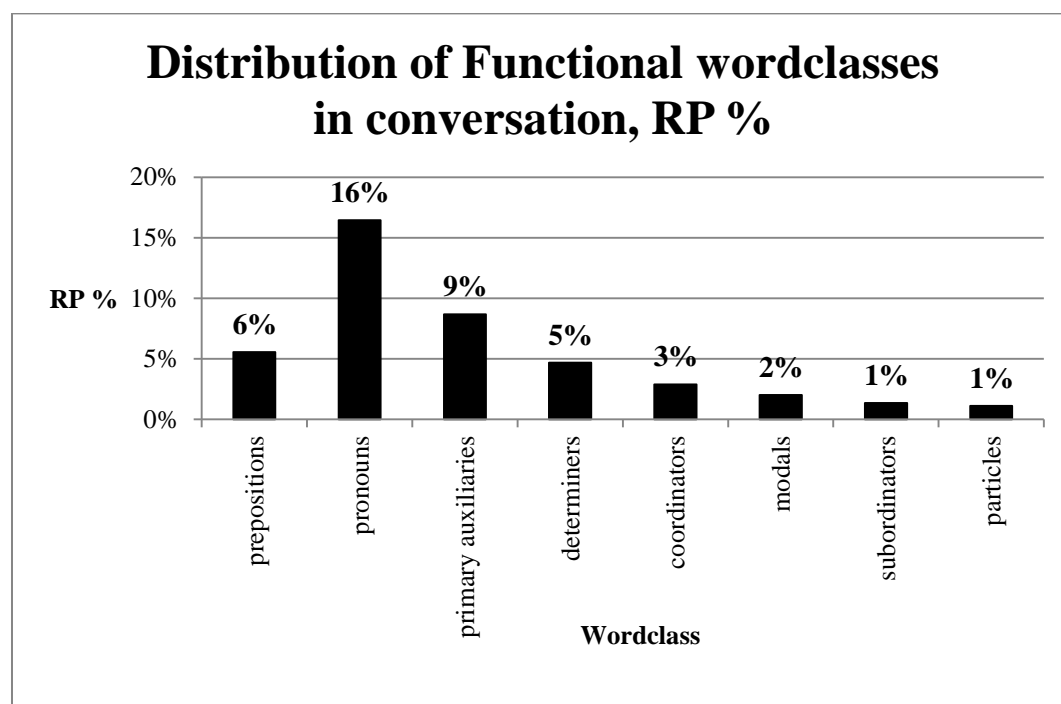
Graph 1 Distribution of Functional wordclasses in Longman corpus (Bieber et al. 1999, p.92)

Distribution of Lexical wordclasses in conversation, RP %

16%	
14%	14%
12%	12%
10%	
RP % 8%	
6% 5%	
4% 3%	
2%	
0%	
adverbs adjectives verbs nouns	
Wordclass	

Graph 2 Distribution of Lexical wordclasses in Longman corpus (Bieber et al. 1999, p.65)

Distribution of Inserts in conversation, RP %

10%	
	6%
RP % 5%	
0%	
Inserts	
Wordclass	
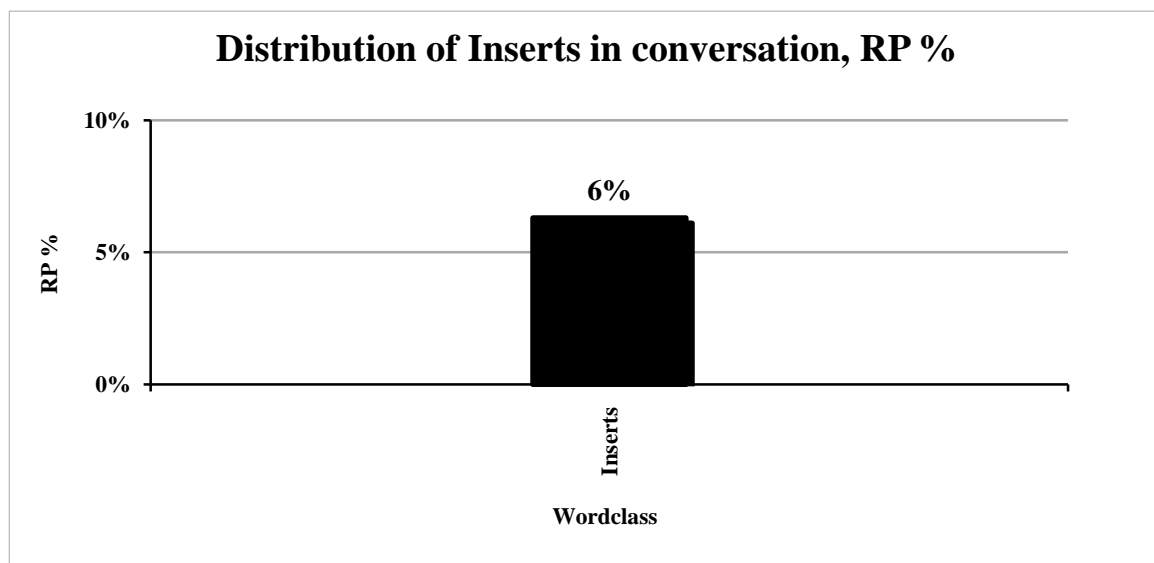
Graph 3 Distribution of inserts in Longman corpus (Bieber et al. 1999, pp.1095-1099)

Table 16. Comparison of five most frequent categories in the Longman's conversation corpus (Biber et al. 1999, p.65;p.92;pp.1095-1099) with findings from participants of this paper

Categories	LGSWE	University graduates					High school graduates				
		U1	U2	U3	Total		H1	H2	H3	Total	
	RP %	RP %	RP %	RP %	Mean RP %	Std dev	RP %	RP %	RP %	Mean RP %	Std dev
Pronouns	16	12.3	12.7	11.2	12.1	0.6	14.0	6.6	16.0	12.2	4.0
Nouns	14	15.0	12.0	17.8	14.9	2.4	5.6	20.2	16.0	13.9	6.1
Lexical verbs	12	10.6	7.7	13.9	10.7	2.5	14.4	13.3	16.7	14.8	1.4
Primary auxiliary	9	2.7	1.8	2.4	2.3	0.4	3.2	7.1	4.5	4.9	1.6
Inserts	6	13.6	11.0	14.1	12.9	1.4	28.0	17.5	9.8	18.4	7.5

According to these three graphs the five most frequent categories in LGSWE spoken corpus are: pronouns, nouns, lexical verbs, primary auxiliaries and inserts.

It is difficult to compare these categories in order. Therefore, the comparison I am going to make here in cases where one or some of these categories are among the five most frequent categories used by the subjects of this study without considering the rankings.

Four of these categories, namely, nouns, inserts, pronouns and lexical verbs are among the five top categories in U1, U3, H1 and H3's corpora. Noun, inserts and pronouns are three of U2's most frequent categories. H2 has used nouns, inserts and lexical verbs among top five categories in his speech.

Considering the relative proportions of these five top categories in the Longman spoken corpus and comparing those figures with the subjects' relative proportions reveal some differences.

The findings from Longman show that in their spoken corpus, the relative proportion of pronouns is 16%. The subjects from the university graduates supplied 11.2%, 12.3% and 12.7% pronouns in their speech. The standard deviation in the usage of pronouns between university graduate participants is 0.6% and as a result there is no considerable difference between the individuals in this group. The relative proportion of the usage of pronouns among high school graduates is 6.6%, 14% and 16% with a standard deviation of 4%. Hence there is much more difference within high school graduate participants than university graduates. Considering the mean RP of high school graduates 12.2% and mean RP 12.1% of university graduates, both of these two study groups are almost equally close to Longman.

The relative proportions of nouns for the university graduates are: 12%, 15% and 17.8% whereas the high school graduate group produced 5.6%, 16% and 20.2%. Compared to the relative proportion of 14% in the Longman spoken corpus, there is a difference. If we consider the mean RP, the high school graduates' RP is closer to Longman's than that of the university graduates. However the standard deviation shows that there is much more difference between the participants in the high school graduates than the university graduates. For example, H1 has used the least amount of nouns and H2 has produced the most amount of nouns both in high school graduates and among all the subjects in this study.

"Lexical verbs" with RP of 12% is the third most frequent category in the LGSWE's conversation corpus. The subjects from the university graduate group have provided 7.7%, 10.6% and 13.9% lexical verbs in their corpora. On the other hand, high school graduate participants have a relative proportion of 13.3%, 14.4% and 16.7% lexical verbs in their material. The standard deviation of 2.5 and 1.4 demonstrate that there is more difference among university graduates' usage of this word class than high school graduates. Regarding mean RP of these two groups 10.7% versus 14.8%, the university graduates are closer to Longman's RP than the high school graduates.

"Primary auxiliaries" is the fourth most frequent word class in Longman's spoken corpus. The relative proportion of this category in this corpus is 9%. University graduates have used this category 1.8%, 2.4% and 2.7% in their corpora whereas high school graduates have produced this category 3.2%, 4.5% and 7.1%. The standard deviation of 0.4% and 1.6% exhibit that the difference within university graduates is smaller than the high school graduates. However, the mean RP of these two groups 4.9% and 2.3% shows that the high school graduates' usage of this category is closer to Longman's relative proportion than the other group.

The fifth most frequent word class in LGSWE conversation corpus is inserts with a relative proportion of 6%. University graduates supplied this word class 11%, 13.6% and 14.1% with a standard deviation of 1.4%. The high school group of this study produced this category 9.8%, 17.5% and 28%. The standard deviation for the high school graduate group is 7.5% which is five times more than the first group. Comparing the mean RP of these two groups 18.4% and 12.9% with Longman's relative proportion of 6% shows that university graduate participants' insert production was closer to Longman's.

3.2.1.5.2 Distribution of lexical, functional words and inserts in Longman corpus

Lexical words presented in graph 2 make up 34% of the words in the Longman spoken corpus. Inserts make up 6% of the words in conversation. As I mentioned earlier the relative proportion distribution of all the function words is not presented in Longman. However, one can add the amount of lexical words and inserts, 40%, and the remainder which constitutes 60% is functional words. So as expected, functional words are much more frequent than lexical words in the spoken language.

According to table 10 in the results section, four subjects: U1, U2, H2 and H3 used function words more than the other two word class types. Content words were among the second word class types which means that Inserts were among the least frequent word class types by these four participants. On the other hand, U3 produced 2% more lexical words than function words. That makes lexical words his most frequent and inserts his least frequent word class types and function words are in the second place for this subject. H1 supplied 46.5% function words. However, what is unexpected is that the second most frequent word class is inserts and not content words in this subject's corpus. I believe that H1's usage of inserts was more unpredicted and strange than U3's. However, four of the subjects from this study, two from each group, have the same rank of these three word classes as Longman's spoken corpus. As mentioned above, the difference lies between two individuals, U3 and H1, rather than groups in this analysis.

3.2.1.5.3 Sentence length

Considering the subjects' mean length of the sentences, the three university graduates are among the top four.

3.2.2 Summary of Linguistic error analysis

Table 17. The usage of linguistic errors made from each participant is presented here

	Linguistic errors															
	University graduates								High school graduates							
	U1		U2		U3		Total			H1		H2		H3		Total		
Total number of words	822		1014		1630		3466			465		504		468		1437		
Areas	N	RP%	N	RP%	N	RP%	N	Average RP%	Std dev	N	RP%	N	RP%	N	RP%	N	Average RP%	Std dev
Semantic errors	10	1.2 %	11	1.1 %	17	1.0 %	38	1.1 %	0.1 %	7	1.5 %	34	6.7 %	43	9.2 %	82	5.8 %	3.2 %
Neologistic error	1	0.1 %	3	0.3 %	15	0.9 %	19	0.4 %	0.3 %	3	0.6 %	2	0.4 %	0	0.0 %	5	0.3 %	0.3 %
Phonological error	4	0.5 %	0	0.0 %	16	1.0 %	20	0.5 %	0.4 %	2	0.4 %	4	0.8 %	19	4.1 %	25	1.8 %	1.6 %
Morphological error	1	0.1 %	6	0.6 %	8	0.5 %	15	0.4 %	0.2 %	2	0.4 %	2	0.4 %	10	2.1 %	14	1.0 %	0.8 %
Total	16	1.9 %	20	2.0 %	56	3.4 %	92	2.5 %	0.7 %	12	2.6 %	42	8.3 %	72	15.4 %	126	8.8 %	5.2 %

3.2.2.1 Summary of morphological analysis

3.2.2.1.1 Comparing the distribution of inflectional morphemes and morphological errors within subjects and groups

The most frequent morpheme used by university graduates is regular plural nouns with 63 instances. This is the second most frequent morpheme produced by high school graduates with 28 cases. The second top frequent morpheme by university graduates is *ing* markers with 60 examples. However, *ing* markers are the most frequent inflectional morpheme in high school graduates' corpora. The *ing* marker was among one of the top two inflectional morpheme used by all subjects which represents how often they have used progressive forms in their speech.

One type of inflectional morpheme which is not used by any of the high school graduates is possessive marker. Two of the university graduates, U1 and U3, have produced it. Verb present morpheme was also found in only one of the high school graduate's corpus, H2.

However, all the university graduates in this paper have used this morpheme. Irregular plural noun morpheme was also used by only one of the high school graduates, H1, while this morpheme was produced by all university graduates.

Considering the relative distribution of inflectional morphemes to total number of words used by the subjects, H1 has produced the least inflectional morphemes both in this study and among high school graduate group. U2 has produced the least morphemes among university graduates. The most frequent morpheme usage was found in H3's corpus.

Table 12 in the previous section indicates that university graduates have used more than twice the number of inflectional morphemes in their speech than high school graduates. However the number of inflectional morphemes in each group should be considered with total number of words produced by that group. Therefore, relative proportion of each inflectional morpheme from each group has been calculated, university graduates had 7.27% and high school graduates had 8.14%. These figures demonstrate that high school graduates used 0.87% more inflectional morphemes than the other group in their interviews. This finding raises the question whether the university graduates who produced more words also made more morphological errors than the high school graduates or the distinction between the relative proportions of these two groups is only based on what type of words they have used. Therefore, the morphological errors made by each participant were studied and compared both within the participants in each group and between the two groups. The average relative proportion of morphological errors made by each group indicates that the university graduates produced 0.4% morphological errors whereas high school graduates had 1% morphological errors in their material. These figures show that high school graduates had two and a half times more morphological errors in their corpora than university graduates. Therefore, even though the high school graduates had a higher relative distribution of inflectional morphemes, they produced more errors of this type. The standard deviation shows that there is more difference within high school graduates concerning this kind of error than between the subjects in the university graduate group.

3.2.2.2 Summary of Semantic analysis

3.2.2.2.1 Comparing the distribution of semantic errors within subjects and groups

The relative proportion of semantic errors within university graduates is between 1% and 1.2%. Therefore the standard deviation of this type of error within university graduates is 0.1%. On the other hand, high school graduates produced semantic errors from 1.5% to 9.2%

having the standard deviation of 3.2%. Therefore, there is more variation within high school graduates than university graduates. H1 has made the least amount of semantic errors within high school graduates. However, all the subjects from the university graduate group have made even less semantic errors than H1.

3.2.2.3 Summary of Phonological analysis

3.2.2.3.1 Comparing the distribution of phonological errors within subjects and groups

The only subject who did not make any phonological errors in this study is the only person with Moderate aphasia U2. U1 had 0.5% and U3 1% of this kind of error in their material making the standard deviation of 0.4% and average relative proportion of 0.5% for this group. The subject with the least amount of phonological error is H1 with only 0.4%. On the other hand, the subject who made most phonological errors in this investigation is H3 with relative proportion of 4.1%. That makes the standard deviation of 1.6% for high school graduates and an average relative proportion of 1.8%.

3.2.2.4 Summary of lexical analysis

3.2.2.4.1 Comparing the distribution of neologistic errors within subjects and groups

H3 is the only subject who did not make any neologistic errors in this research. The participant with only one neologistic error is U1. The standard deviation for university graduates is 0.3% with an average relative proportion of 0.4%. The standard deviation of the high school subjects is the same with an average relative proportion of 0.3%.

3.2.2.4.2 Type Token Ratio Analysis

LGSWE shows that there is a much lower type-token ratio in the conversation than the other registers. The reason that there is not as much variation in the choice of words in conversation as in other registers is that spoken language is produced spontaneously with very little time to plan ahead. According to Longman, type token ratio varies with the number of the words used in conversation. That means there are many more repeated words in longer transcriptions of speech and as a result they have much lower TTR than shorter transcriptions of speech. One of the typical characteristics of spoken language is repetition. It is used either in order to emphasize or to make sure that the message gets across the hearer (Biber et al. 1999, p.53).

Graph 4 has been taken from Longman's spoken corpus findings which are discussed above. The graph shows that the mean TTR for the person who speaks 100 words is 64.44 whereas

the speaker with 500 words has the mean TTR of 37.16 and the expected mean TTR for the speaker who produces 1000 words is 30.

The reader should bear in mind that the figures are the Longman's spoken corpus findings from ordinary people. However, these figures are compared with the TTR of the subjects in this study

As was discussed in the results section two of the participants of this study, U2 and U3, had the lowest TTR 18.06 and 19.33. On the other hand, these two participants used the most number of words. Therefore, I have tried to make an adjustment by taking the number of words produced by each subject and the mean TTR of those number of words in Longman's spoken corpus in order to discover similarities or differences compared to Longman. The line graph demonstrates which mean TTR is expected according to the number of words used. It shows clearly how the mean of TTR decreases with the increase in number of words. The symbols on the graph exhibit each participant's TTR in the data. A comparison with the black line shows how close or far they were from Longman's figures. Table 18 makes it easier and more clear for the reader to compare these numbers. Four different types of information are provided in this table: the subjects' TTR, Longman's TTR at the same number of words as the subjects, the difference between Longman's and the subjects' TTR both as number and percentage. As indicated in the table, the subject with the closest anticipated TTR mean to Longman's is U1 with a difference of 2.28 (7.3%). H3 is the second participant with only 4.27 (11.2%) difference with Longman's spoken corpus. U3 is the third participant with 5.20 (21.2%) difference compared with LGSWE spoken corpus. H1 with 10.87 (28.5%) is the fourth participant and U2 with 10.92 (37.7%) is the fifth closest to Longman's conversation TTR. H2 with 14.96 (40.8%) is the farthest from Longman's spoken corpus TTR.
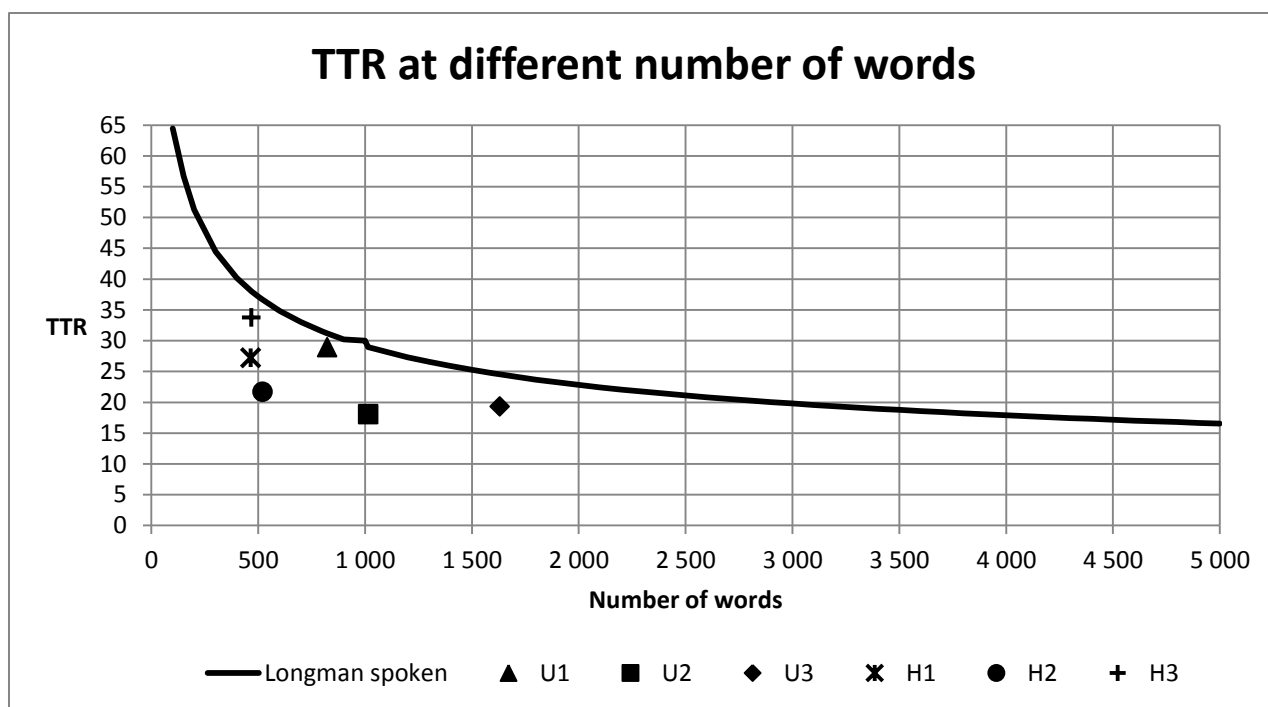
Graph 4 TTR at different number of words in the Longman spoken corpus (Biber et al. 1999, p.53) and TTR for U1-U3 and H1-H3.

Table 18. Comparing TTR of the subjects U1-U3 and H1-H3 with the Longman's spoken corpus findings (Biber et al. 1999, p.53)

	U1	U2	U3	H1	H2	H3
TTR subjects	28.92	18.06	19.33	27.25	21.69	33.76
TTR Longman spoken language at the same number of words as subject	31.20	28.98	24.53	38.12	36.65	38.03
Difference in TTR from Longman spoken language at the same number of words	-2.28	-10.92	-5.20	-10.87	-14.96	-4.27
% diff from Longman spoken language at the same number of words	-7.3%	-37.7%	-21.2%	-28.5%	-40.8%	-11.2%

4 Discussion

In the following, the results of this paper will be discussed in relation to the aim and hypothesis of this study. In addition, the general results of this paper will be compared with five other studies mentioned in the "previous research" section.

This investigation was an attempt to find out if there is a difference between the language of high school and university graduate PWAgr. I think that part of studying at university at higher levels is to read and to some extent produce academic articles. This makes university graduates familiar with formal language and as a result it would automatically affect their language production to spontaneously follow grammatical rules and speak more correctly. However, the type of job a person has after obtaining a degree of higher education is also very important. I think this would be only the case if the person is still expected to read academic literature in order to overcome the challenges that his job throws at him. Hence, the two parameters, educational background and occupation, go together. Therefore, my hypothesis is that there is a difference between the grammatical accuracy of university graduates and high school graduates in general. The research question of this study is if the assumed difference would still be valid after these persons had a stroke.

In order to verify the hypothesis and find an answer to the research question, three high-school graduates PWAgr and three university graduate PWAgr were selected and different types of linguistic analyses have been performed on their spoken language production. The findings of these analyses will be discussed below.

The results from table 9 (p.38) show that university graduates used a greater number of grammatical categories in their corpora than the high school graduates. As displayed in table 11 (p.43), there is also a difference between the number of words used by each group. High school graduates together produced 1437 words whereas university graduates supplied 3466 words. Comparing the only person with Moderate aphasia in this study, U2, with other subjects with Mild aphasia demonstrates that he has used more than double the number of words than each of the high school graduates. The reader should bear in mind that regarding the number of words university graduates used, they had the chance to make more mistakes or leave out the required words more than two and a half times. However, they omitted the required words, less than one per cent on average more than high school graduates. It is the same situation for incorrect words provided by each group. Again, it is less than one fifth of one per cent on average difference between the groups. The reader should take into account

that these figures are only based on the sentences which could be reconstructed and analyzed in this research.

Taking the standard deviation of these two groups for all the five categories in table 16 (p.60) into account, it shows that for the high school graduates there was a greater divergence between each participant in that group than university graduates. However, the high school graduates' mean RP was closer to the Longman's in two of the categories, nouns and primary auxiliaries, and the university graduates' mean RP was closer to the Longman for two other categories, lexical verbs and inserts, than the high school graduates. These two groups' mean RP for the pronouns was almost equally close to The Longman's.

Graphs number one, two and three in the previous section were designed to make it possible to compare the usage of the three word category types: lexical, functional and inserts by the subjects of this study with Longman's conversation corpus findings of frequency of these three main word class types. There was no difference between these two subject groups concerning this analysis. A difference exists between two individuals, one of each subject group, which has been taken up in section, 3.2.1.5.2.

Another kind of linguistic analysis performed on the data was examining semantic errors produced by each participant and comparing the number of errors with other subjects in that group and within these two study groups. Taking the average relative proportion of each group into account, high school graduates made at least five times more semantic errors in their interviews than the other group.

The third type of error investigated in this paper is phonological errors. According to table 17 (p.63), the average relative proportions of this error suggest that high school graduates made at least three times more phonological errors than the other group. Comparing the standard deviation of these two groups for phonological errors shows that the difference within high school graduates is again more than university graduate participants.

The last type of error analysis that was done on the data is neologistic errors. Inspection of Table 17 divulges that on average university graduates made almost 30% more neologistic errors than the other group. The standard deviation of this part shows that there is as little difference between individuals in the high school group as in the university graduate subjects.

The WordSmith analysis revealed some information about TTR and sentence length of each participant's text. The type-token ratio of this study's participants was then compared with

Longman's spoken corpus in Table 18 (p.67). According to what is discussed in the Type Token Ratio Section 3.1.5.1.1., we can conclude that although the TTR from the subjects' data in the result indicated that the high school graduates had higher TTR than the university graduates, after considering the amount of words each participant had produced and comparing that with the expected TTR from Longman's spoken corpus one can draw the conclusion that university graduates were also a little closer to what was anticipated. However, the university graduates used more words and as a result had more repetitive words and fewer distinctive (types) words than high school graduates.

Regarding sentence length, table 11 shows that university graduates had a mean of 5.55 in their sentences compared with 3.88 for the high school graduates. However, there are more group internal differences between high school graduates than among university graduates.

To summarize, university graduates used a greater number of word classes and produced two and a half times more words in their materials than high school graduates. Regarding the error analyses of the data, high school graduates made two and a half times more morphological errors, three times more phonological errors and five times more semantic errors than university graduates. And the university graduates made around thirty per cent more neologistic errors than high school graduates. This means that high school graduates made considerably more errors in three out of the four linguistic phenomena discussed above. The standard deviation of these two groups in table 17 indicates that in four of five error analyses, there is more difference among high school graduate participants than university graduate subjects. In one case, neologistic errors, the two groups had the same standard deviation.

Having one person with Moderate agrammatic aphasia among university graduates strengthens the findings of this study. The hypothesis holds true as regards difference between language production of university and high school graduate PWAgr of this research. The former group has produced more words, more word classes and made fewer errors.

The findings of this study are consistent with Fernandez (2011) findings mentioned in the "Previous Research" Section, 1.1. I also found that educational background has an effect on the accuracy of the subjects' language performance. Tainturier, Tremblay and Lecours (1992) and Béland and Lecours (1990) also concluded in their research that there were some differences between the language performances of healthy adults regarding their educational length. However, their subjects, unlike the participants of the present paper, were healthy adults.

The main outcomes of studies carried out by Soares and Ortiz (2008) and Connor et al. (2001) are inconsistent with the results of this study. The researchers did not find any association between educational level and better language performance in these two investigations.

The results of the investigation of three high school and three university educated persons with agrammatic aphasia suggest that there is a difference between the spoken language production of these two study groups. There is no earlier research, to my knowledge, with the same focus of the impression of educational length and occupation status on the linguistic performance of PWAgr. Therefore, the comparisons made above are only based on the general findings of other studies and the researchers' suggestions on the influence of education on the language performance of their participants. The reader should bear in mind that the comparisons are made indirectly due to the differences in the studies' methods.

5 Conclusion and Future Directions

The goal of this paper is to investigate whether there is a difference between the spoken language production of three American high school graduate participants who had physically challenging jobs before their stroke and three American university graduate subjects who had mentally challenging jobs before their illness. The six subjects were clinically categorized as having non-fluent Broca's aphasia.

The findings of this study suggest that there is a difference between the spoken language performances of these two groups. The high school graduates used fewer words, fewer grammatical categories and had many more linguistic errors in their speech than university graduate participants.

The reader should keep in mind that this is a small study and more investigation on a bigger scale is necessary in order to make any generalizations. The difference found in this research is valid considering these six subjects.

Future research should focus on a greater number of participants. The researcher can try to find subjects with the same variables. For example the age at onset and testing as well as aphasia duration vary between participants of this study. In addition, it was not possible to separate the "level of education" and "multilingualism" factors in U1 in this study. However, the effect of the multilingualism on the spoken language performance can be treated separately in the future investigations. The subjects' written language as well as gestures they use in order to express themselves could also be taken into account. Furthermore, a cross-linguistic study can be done in order to find out if the findings of this paper are applicable in different languages. If future researchers also find a difference then it may be fruitful for speech pathologists, during the rehabilitation process, to pay even more attention to the educational length and the kind of job the patients had before their illness.

6   Appendix

6.1   Explanation of the tables 3-8

6.1.1   Detailed explanation of the results from table 3 (p.23)

ADJECTIVES

U1 used 31 adjectives which constitute 3.8% of the total words.

ADVERBS

The subject produced 25 adverbs which make up 3% of the total number of words. In addition, U1 used 2 quantifiers and one *wh*-word as adverbs. She also used one adverb of frequency. All adverbs produced by U1 comprise 3.4% of all the words.

One adverb *there* was missing in U1's corpus which constitutes 3.8% of total reconstructed adverbs used by this subject. In other words, 96.2% of other adverbs analyzed here were used correctly.

ADVERBIAL PARTICLES

There are nine adverbial particles in U1's corpus which represent 1.1% of the total words. This is the third least frequent category in the corpus.

CONJUNCTIONS

Two different types of conjunctions, coordinating conjunctions and subordinating conjunctions, were used by this subject. Coordinating conjunctions were used in 87 instances, comprising 10.6% of total number of words. Twelve subordinating conjunctions were used which constitute 1.5% of all words. U1 used 99 conjunctions in her speech. Both conjunctions together represent 12.1% of all the words. This is the fourth most frequent category in this corpus.

Two coordinating conjunctions *so* and *and* were missing in the material. Only 2.2% of coordinating conjunctions were missing when required.

COPULAS

Copulas were used 33 times. This category constitutes 4% of the total number of words.

The copular verb *is* was left out two times when needed. Three copular verbs were used in a wrong tense. Two times instead of the past tense *was* the present tense *are* was provided and once the present form *was* instead of *is* was used.

ARTICLES

U1 used both definite and indefinite articles. Definite articles were used in 50 cases and represent 6.1% of total number of words, whereas indefinite articles were used 12 times which make up 1.5% of all the words. 62 articles constitute 7.6% of total number of words.

In 12 instances, the definite article *the* and in two cases the indefinite article *a* were missing. In other words, 19.4% of definite articles and 14.3% of indefinite articles were not used by this subject when it was obligatory in the sentences. In another example, the indefinite article *a* was used instead of the definite article *the*.

DETERMINERS

Demonstrative determiners such as *this* and *that* were found five times in U1's speech. Possessive determiners such as *my,* were used in eight cases by U1. Numerals were used as determiners in 16 instances. Quantifiers were also used as determiners five times by this subject.

All 34 determiners together constitute 4.1% of total number of words.

Quantifier *some* was missing in two cases where it should have been used as a determiner.

NUMERALS

There were no other numerals in the corpus.

EXISTENTIALS

Existential *there* occurred once among the words this participant supplied.

INFINITIVE MARKERS

Infinitive marker *to,* was used five times.

INSERTS

Five different types of inserts, i.e. discourse marker, farewell, interjections, response and hesitators were used by this subject. This category was used 112 times in U1's speech comprising 13.6% of total number of words.

PRIMARY AUXILIARIES

Primary auxiliaries occurred 22 times in U1's speech. This represents 2.7% of all the words.

MODAL AUXILIARIES

The subject produced no modal auxiliaries.

NEGATORS

Not was used 21 times in U1's speech which represents 2.6% of the total words.

NOUNS

U1 produced both common and proper nouns. Common nouns were found in 116 instances which constitute 14.1% of total number of words. Of those 116 cases, 102 were singular nouns, 13 regular plural nouns and one irregular plural noun.

Proper Nouns such as *Cinderella* were used seven times by this participant.

Four required common nouns were left out in the context.

In one of the sentences the proper noun *Cinderella* was not utilized though it was necessary to use in that context.

11.1% of omissions in this participant's material were common nouns.

PRONOUNS

Two types of personal pronouns, nominatives and accusatives, were used in U1's speech during the interview. Nominative personal pronouns were produced 63 times and accusative personal pronouns were supplied in 14 instances comprising 7.7% and 1.7% of total number of words respectively. Demonstrative pronouns were found ten times (1.2%) and indefinite pronouns were used in nine cases comprising 1.1% of total number of words. Relative

pronouns were also used five times. In general, all different types of pronouns together constitute 12.3% of all the words.

Three personal pronouns *he* and one personal pronoun *she* were missing from U1's corpus. Four omissions of nominative personal pronouns represent 11.1% of all the missing words in this subject's context.

PREPOSITIONS

The subject used prepositions 53 times comprising 6.4% of the total words.

Preposition *for* in three examples and preposition *by* in one instance were left out from the context. Four required prepositions were missing in this participant's material forming 11.1% of total omission words.

POSSESSIVE CLITICS

One possessive clitic was used by U1. This category along with existentials is the least frequent categories here.

VERBS

U1 used 86 lexical verbs in her speech. Verbs constitute 10.5% of total number of words in this material.

Two lexical verbs, *happened* and *use* are missing from the material. In addition, three lexical verbs produced by this subject have the wrong tense. Once the present form of the verb *go* was required instead of the simple past tense. Another time it was the opposite. The past tense *went* was needed instead of present form *go* and once U1 used progressive form *dancing* instead of the simple present form of this verb *dance*.

OTHER COMMENTS

One of the sentences produced by U1 is *oh and I no and and I uh was um just recently m computer um special for uh this*. The reconstructed form of this sentence is *and I just recently use special computer for this*. According to the reconstructed form, there is an adjective-noun inversion in this example.

Another sentence provided by this subject is *I'm fan no of Cinderella in the past*. She probably wanted to say *I was no fan of Cinderella in the past*. So she has changed the place of the noun *fan* and the quantifier *no*.

6.1.2 Detailed explanation of the results from table 4 (p.26)

ADJECTIVES

U2 used 60 adjectives forming 5.9% of total number of words.

ADVERBS

130 instances of adverbs were found. In addition, 11 quantifiers and seven *wh*-words were used as adverbs in the corpus. No adverb of frequency was used. All adverbs together constitute 15.9 % of total words used by this subject. One adverb of frequency *sometimes* as well as another adverb *back* were left out from the required context.

ADVERBIAL PARTICLES

Adverbial particles were produced in 13 cases comprising 1.3% of total number of words.

CONJUNCTIONS

Ninety-seven examples of coordinating conjunctions and three instances of subordinating conjunctions were found in the corpus. Coordinating conjunction *and* was missing three times from the corpus.

COPULAS

This category was used 53 times during U2's speech. It represents 5.2% of the total number of words.

The copular verbs *am*, *are* and *were*, were missing once each from the material. Fourteen copular verbs *is,* as well as six copular verbs *was*, were also left out when required from this corpus. The subject has omitted 30.3% (23 instances) of the copular verbs where it was necessary.

One incorrect tense of the copular was produced by U2. He used copular *is* instead of *was* in the following sentence *and then uh maybe died sometimes it's easy sometimes difficult*.

ARTICLES

Thirty-three examples of definite articles and 15 instances of indefinite articles were used. These two types of articles together make up 4.8% of the total words. Indefinite article *a* was omitted in eight cases and was used once incorrectly instead of the definite article *the* in the material.

Definite article *the* was eliminated in 12 instances.

26.7% of definite articles and 34.8% of indefinite articles were omitted in the obligatory contexts.

DETERMINERS

Seven demonstrative determiners and three possessive determiners were found in the corpus. Furthermore, 32 cases of numerals and 19 quantifiers were used as determiners. U2 also used *either* as a determiner once. All different forms of determiners constitute 6.2% of total words used in this corpus.

The quantifier *some* was excluded once when it should have been used as a determiner. In addition, possessive determiner *my*, was eliminated four times (57.1%) in the obligatory context.

NUMERALS

There were 10 numerals in the material.

EXISTENTIALS

Existential *there* was used three times. Once *there* was not used where it was required.

INFINITIVE MARKERS

This category was used in 12 cases.

INSERTS

112 examples of inserts: discourse marker, expletives, farewell, hesitators, interjections, polite speech acts and responses were produced. *Yes* was omitted once in an obligatory context.

PRIMARY AUXILIARIES

U2 used 18 primary auxiliaries.

MODAL AUXILIARIES

Modal auxiliary *can* was produced in the following sentence *so uh maybe many many years ago I can't talk*. However, as he used *many years ago*, this should have been in the past tense *could* and not in the present form.

NEGATORS

The subject produced 29 negators.

NOUNS

The participant used 110 common nouns and 12 proper nouns. Of those 110, they were 74 singular nouns, 31 regular plural nouns and five cases of irregular plural nouns. All different forms of nouns together make up 12% of total number of words. Four common nouns *man*, *people*, *business* and *boy* were eliminated from the corpus. In other words, 3.5% of obligatory common nouns were left out in the context.

PRONOUNS

There were 35 demonstrative pronouns, 16 indefinite pronouns, 18 accusative personal pronouns and 59 nominative personal pronouns in the corpus. No interrogative pronouns or relative pronouns were found.

All these different pronouns together form 12.7% of all the words used.

Eighteen nominative personal pronoun *it*, seven *I*, one *he* and one personal pronoun *she* were left out from the material.

Accusative personal pronoun *it* was also required in the following sentence *I eat* however, it was not provided.

PREPOSITIONS

Sixteen instances of prepositions were used.

POSSESSIVE CLITICS

No possessive clitics were found.

VERBS

Seventy-eight examples of lexical verbs were produced by this subject. Nine *hads*, one *say*, one *improving* and one *think* were excluded from this corpus. 12.1% of omissions in this corpus belong to lexical verbs.

OTHER COMMENTS

Two different forms of inversion occurred in U2's material.

In the following sentence *you know is it coming pretty good you know.* There is a subject-verb inversion. It should be a declarative sentence not an interrogative. Therefore, he should have changed the place of *it* and *is*.

In this example *and then the man is the two people uh small something to do with it*. There is an adjective-noun inversion considering the reconstructed form *two small people had something to do with it.*

6.1.3 Detailed explanation of the results from table 5 (p.29)

ADJECTIVES

U3 used 19 adjectives.

ADVERBS

There were 185 adverbs in the corpus. In addition, the subject provided one quantifier and one *wh*-word as adverb. No adverb of frequency was found. Two adverbs *ago* and *maybe* are excluded from obligatory context in U3's corpus.

ADVERBIAL PARTICLES

Twenty-nine adverbial particles were found in this corpus. The adverbial particle *up* was missing from this subject's speech in one example.

CONJUNCTIONS

Seventy-three examples of coordinating conjunctions and two instances of subordinating conjunctions were found in the corpus. Conjunctions make up 4.6% of total words.

Coordinating conjunction *and* was omitted in one of the instances when required.

COPULAS

There were nine copulas constituting 0.6% of all the words in the corpus. Six examples of *was*, 35 instances of *is* and one case of *are* are left out in the obligatory context here. 25% of the missing words in this material belong to copular verbs and 82.4% of required copular verbs were missing here.

ARTICLES

There were 92 definite articles and eight indefinite articles in the material. Articles make up 6.1% of the total words. Definite article *the,* was eliminated in 58 cases where it was obligatory to use it. Once it was also produced where it was not needed. Indefinite article *a* was also missing in 12 compulsory cases. 38.7% of the definite articles were missing in the required context in U3's speech and the indefinite article was omitted in 60% of the obligatory cases.

DETERMINERS

The subject produced nine demonstrative determiners and one possessive determiner. Moreover, 31 numerals, eight quantifiers and one *wh*-word were used as determiners. All these different types of determiners represent 3.2% of all words used.

NUMERALS

Eighty-one examples of this category were found in the corpus.

EXISTENTIALS

Only one existential was produced. Existential *there* was once omitted in the required sentence.

INFINITIVE MARKERS

Only one instance of this category was found. Infinitive markers together with existentials account as the least frequent categories in this corpus. Infinitive marker *to* was eliminated once in the obligatory context.

INSERTS

230 cases of inserts: discourse markers, farewell, interjections, responses and hesitators, were found in the corpus.

PRIMARY AUXILIARIES

U3 used 39 primary auxiliaries. Primary auxiliary *did* was eliminated before not from the following sentence *maybe not fall but slipped*.

MODAL AUXILIARIES

There were no modal auxiliaries here. However, one modal auxiliary was needed in one case but it was not provided.

NEGATORS

The subject used negator *not* 62 times. In this case, *then well not bicycle on the ground* the negator *not* should not have been used as there is a bicycle on the ground. So, the usage of this category is not correct here.

NOUNS

The participant used 278 instances of common nouns and 12 examples of proper nouns. Of those 278, there were 253 singular nouns, 21 regular plural nouns and 4 irregular plural nouns.

One common noun *window* is missing after *the* in the following sentence *fourth picture man looking out the*.

PRONOUNS

There were 19 indefinite pronouns, five accusative personal pronouns, 153 nominative personal pronouns and three interrogative pronouns in U3's material. The subject used only one quantifier as a pronoun. No demonstrative or relative pronouns were found.

Twenty nominative personal pronouns and one accusative personal pronoun were omitted in the obligatory contexts. Accusative personal pronoun *it* was produced in two cases where it was not needed.

PREPOSITIONS

U3 produced 47 instances of prepositions.

POSSESSIVE CLITICS

Two possessive clitics were used.

VERBS

Lexical verbs were used 227 times, comprising 13.9% of all the words in the corpus. Seven lexical verbs were missing from the corpus. In addition, the subject used the wrong tense for six of the lexical verbs in the material. Past tense of the following verbs *take*, *buy* and *rain* were required instead of the present forms. In two cases the progressive forms *throwing* and *shattering* were used where the past tenses *threw* and *shattered* were needed. In one instance the subject produced *goed* instead of the irregular form *went*.

OTHER COMMENTS

In the following example *maybe seventy five pe rcent uh improve it*, there is a subject-verb inversion as the reconstructed form is *it has improved*.

In another instance, there is again a subject-verb inversion. The subject said *then prince um well maid prince well marries* the reconstructed form is *the prince marries the maid*. The word *maid* should be moved. It is an object and it should appear after the verb *marries*.

In this case *then I think so maybe Ju:ne fifteenth about* the participant should have used the preposition *about* before *June fifteenth* not after.

6.1.4 Detailed explanation of the results from table 6 (p.32)

ADJECTIVES

The subject produced 15 adjectives.

ADVERBS

Ten adverbs were found. H1 also used one quantifier and four *wh*-words as adverbs in his speech. No adverb of frequency was produced.

ADVERBIAL PARTICLES

Four examples of adverbial particles were used.

CONJUNCTIONS

Twenty four coordinating conjunctions and one subordinating conjunction were found in the corpus making this category up to 5.4% of all the words.

COPULAS

H1 used 18 copulas. Seven copular verbs *is* and one *be* are missing in the material. In one case, present copular verb *is,* was produced instead of past tense *was*.

ARTICLES

Four definite articles and four indefinite articles were found. In eight cases definite article *the* and in one instance the indefinite article *a* were not provided in the required contexts.

DETERMINERS

One demonstrative determiner and two possessive determiners were found. The subject used one quantifier as a determiner. No other determiners were found. Determiners together with adverbial particles were the second least frequent categories in the corpus.

Possessive determiner was missing in one example.

NUMERALS

There were 22 numerals in the material.

EXISTENTIALS

No existential *there* was found.

INFINITIVE MARKERS

Six instances of infinitive markers were produced.

INSERTS

131 inserts were found including: discourse markers, expletives, greetings, hesitators, interjections, polite speech acts and responses were found. This makes up to 28.2% of total words in this corpus.

PRIMARY AUXILIARIES

Fifteen primary auxiliaries were produced.

MODAL AUXILIARIES

There were 18 modal auxiliaries in the material.

NEGATORS

The subject used 24 negators in his speech comprising 5.2% of the total words.

NOUNS

There were 22 common nouns and four proper nouns in the corpus. Out of 22 common nouns, 17 singular nouns, four regular plural nouns and one irregular plural noun were found.

One irregular plural noun *children*, two regular plural nouns *boys* and three singular nouns *ball* were not provided in the context when necessary. 14.6% of the omissions in this corpus belong to common nouns.

PRONOUNS

Six demonstrative pronouns, four accusative personal pronouns, 49 nominative personal pronouns and six interrogative pronouns were used. There were neither indefinite pronouns nor relative pronouns in the corpus. Pronouns contain 14% of total number of words.

Four examples of nominative personal pronouns *I*, eight *he* and three *it* were not supplied in the required sentences making it 36.6% of omission in this corpus.

One instance of accusative personal pronoun *it* was also missing in the obligatory context.

PREPOSITIONS

The subject used two prepositions.

POSSESSIVE CLITICS

There are no possessive clitics.

VERBS

H1 produced 67 lexical verbs comprising 14.4% of total words. One lexical verb was missing in this material. In one case, past tense *made* is used instead of present form *make*.

6.1.5 Detailed explanation of the results from table 7 (p.34)

ADJECTIVES

H2 used three adjectives.

ADVERBS

The subject produced four adverbs. He used no adverb of frequency. He did not use quantifiers or *wh*-words as adverbs.

ADVERBIAL PARTICLES

Three adverbial particles were found in the corpus. Adverbial particles and adjectives together are the second smallest categories here.

CONJUNCTIONS

There were 12 coordinating conjunctions and no subordinating conjunctions in the material.

COPULAS

Thirteen examples of copulas were produced comprising 2.6% of total words. The copular verb *is* was missing once in the subject's material.

ARTICLES

Eighty-five instances of definite article and six cases of indefinite article were used by the subject making this category up to 18.1% of total number of words. Nine examples (33.3%) of the required definite articles were missing in this participant's material. Two indefinite articles were also not provided by this subject. In this instance *they're um the mother is um getting the uh a umbrella to the kid*. Indefinite article *an* is required instead of *a* before *umbrella*.

DETERMINERS

H2 used three possessive determiners and eight numerals as determiners in his speech. There were no demonstrative determiners or other forms of determiners in the corpus.

NUMERALS

There were no numerals in the corpus.

EXISTENTIALS

The subject produced one existential.

INFINITIVE MARKERS

No infinitive marker was found in the corpus. However, one was required but the subject did not supply it.

INSERTS

Three different types of inserts: interjections, response and hesitators were found. The participant used 88 inserts comprising 17.5% of all the words.

PRIMARY AUXILIARIES

Thirty six examples of modals make up 7.1% of the total words.

MODAL AUXILIARIES

No modal auxiliaries were produced.

NEGATORS

There were 16 examples of negators making up to 3.2% of the total words.

NOUNS

101 instances of common nouns and only one proper noun were produced by the subject which make this category up to 20.2% of in the corpus.

The following common nouns *event*, *ball*, *umbrella*, *river* and *girl* were missing in H2's corpus. The common noun *boy* was needed twice but it was not supplied in the obligatory contexts.

PRONOUNS

H2 produced four accusative personal pronouns, 27 nominative personal pronouns and only two examples of interrogative pronouns. He did not use any demonstrative, indefinite or relative pronouns. H2 did not even use any quantifiers as pronouns. All the pronouns together make up 6.6% of the total number of words. In one example, nominative personal pronoun *she* was missing.

PREPOSITIONS

There were 24 prepositions in the corpus. This category constitutes 4.8% of all the words. Once the preposition *about* and twice the preposition *in* were not produced in the obligatory text.

POSSESSIVE CLITICS

There were no possessive clitics in the material.

VERBS

Sixty-seven lexical verbs were found in the corpus. Three lexical verbs were missing in the compulsory contexts. In one example, past tense of the lexical verb *understand* was supplied instead of the present form. So, the subject used the wrong tense.

6.1.6 Detailed explanation of the results from table 8 (p.36)

ADJECTIVES

There were 14 adjectives in this corpus.

ADVERBS

Eight examples of adverbs were found in the material. The subject used two *wh*-words as an adverb. There was no adverb of frequencies. The adverb *ever after* was missing in the material.

ADVERBIAL PARTICLES

H3 used five adverbial particles here.

One example of adverbial particle *up* was left out in the required context.

CONJUNCTIONS

Thirty-two instances of coordinating conjunctions and four examples of subordinating conjunctions were produced by this subject. This category comprises 7.7% of all the words in this subject's material.

COPULAS

Thirteen copulas were supplied by the subject. Four cases of copular verb *was* and three examples of copular *is,* were not found in the obligatory context. 19.4% of omitted words in the corpus belong to this category.

ARTICLES

Forty-five definite articles and seven indefinite articles were produced by the subject.

In five cases, the definite article *the* was not used in the required contexts. One indefinite article *a* was missing in the corpus. In addition, the indefinite article *a* was produced in one example which was not required.

DETERMINERS

Three possessive determiners and only one demonstrative determiner were found in the corpus. The subject also used seven numerals as determiners. This category makes up to 2.3% of the total number words.

NUMERALS

Seven numerals were found in the material.

EXISTENTIALS

There was no existential *there* here.

INFINITIVE MARKERS

There are 11 cases of infinitive markers in the corpus.

INSERTS

Forty six inserts that the participant used were among hesitators, interjections, polite speech acts and response.

PRIMARY AUXILIARIES

Twenty one primary auxiliaries were found here making up to 4.5% of all the words.

MODAL AUXILIARIES

H3 used only one modal auxiliary during her speech.

NEGATORS

Four instances of negators were produced.

NOUNS

The subject used 65 common nouns and 10 proper nouns. Among the common nouns there were 51 singular nouns and 14 regular plural nouns.

One common noun *leg* was not produced in the required sentence.

PRONOUNS

H3 used seven accusative personal pronouns, 62 nominative personal pronouns, one interrogative pronoun and one relative pronoun. She also used four quantifiers as pronouns. There were no demonstrative or indefinite pronouns. All the pronouns together make up 16% of the total number of words and the second most frequent category together with nouns in this material.

Four examples of nominative personal pronoun *it*, six cases of *I*, one instance of *she* and one *they* were not provided by the subject in this corpus. These twelve omitted nominative personal pronouns make up 33.3% of all the omitted words in H3's corpus. In addition, one example of accusative personal pronoun *it* was left out from this corpus.

PREPOSITIONS

There were 10 prepositions in the corpus. One example of the preposition *at* was not found in the compulsory context.

POSSESSIVE CLITICS

No possessive clitics were used here.

VERBS

Seventy-seven instances of lexical verbs make up 16.5% of all the words used here. Six lexical verbs *got*, *had*, *talked*, *move*, *went* and *lived* were not supplied by this subject where needed.

In one example *got* was used instead of present form *get*.

6.2 Some examples of conversation

6.2.1 An example of conversational extract from the Longman's spoken corpus (Biber et al. 1999, p.1040)

D1: Mom, I, give me a rest, give it a rest. I didn't think about you. I mean, I would rather do it. <unclear> some other instance in my mind.

P1: Yeah, well I can understand you know, I mean [unclear] Hi I'm David's mother, try to ignore me.

D2: I went with a girl like you once. Let's serve this damn chilli.

M1: Okay, let's serve the chilli. Are you serving or not dad?

J1: Doesn't matter.

P2: Would you get those chips in there. Michael, could you put them with the crackers.

J2: Here, I'll come and serve it honey if you want me to.

P3: Oh wait, we still have quite a few.

D3: I don't see any others.

6.2.2 An example of conversational extract from U1's interview (Talkbank, 2012g)

Interviewer: Well how (a)bout your first memories after the stroke?

Interviewer: Can you what can you tell me about that?

U1: Nothing.

U1: I woke up when I was taken uh out of +...

U1: Well I, I was in the hospital for an operation of the heart.

U1: Oh.

U1: And it's gunning alright.

U1: And one day afterwards my stroke.

U1: So I have nothing until uh um I woke up uh [x 5] in the middle of taking me to uh another hospital.

Interviewer: Did you know what was happening?

U1: No

U1: But the hospital told me.

U1: The ha um the hospital told me but I don't remember.

6.2.3 An example of conversational extract from H2's interview (Talkbank, 2012h)

Interviewer: Do you remember when you had your stroke?
H2: No.
Interviewer: What have you, what are your first memories after the stroke?
H2: H hospital.
Interviewer: Can you say more about it?
H2: Um hm +...
H2: Um um hospital and um I'm goin(g) uh I don't I don't understand.
H2: I don't know.
Interviewer: How (a)bout your recovery?
Interviewer: Can you tell me what you've done since you had your stroke to try to get better?
H2: Um Denise hu helped me.
Interviewer: Has she seen you right from the beginning?
H2: Yes.
Interviewer: How long ago was your stroke?
H2: Hm th four years ago.
Interviewer: So you've been seeing her for four years.
H2: Yes.

References

Ahlsén, E., 2006. *Introduction to neurolinguistics*. Amsterdam: John Benjamins Publishing Company.

Albert, M.L., Goodglass, H., Helm, N.A., Rubens, A.B. and Alexander, M.P., 1981. *Clinical Aspects of Dysphasia*. Vienna and New York: Springer.

Avrutin, S., 2001. Linguistics and agrammatism. *GLOT International,* 5(3), pp.87-97.

Bastiaanse, R., and Thompson, C.K., 2003. Verb and auxiliary movement in agrammatic Broca's aphasia. *Journal of brain and language*, 84(2), pp.286-305.

Béland, R., and Lecours, A.R., 1990. The mt-86 β aphasia battery: A subset of normative data in relation to age and level of school education. *Aphasiology*, 4(5), pp.439-462.

Berndt, R.S., and Caramazza, A., 1980. A redefinition of the syndrome of Broca's aphasia: Implications for a neuropsychological model of language. *Journal of Applied Psycholinguistics*, 1(3), pp.225-278.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E., 1999. *Longman Grammar of Spoken and Written English.* Essex: England: Longman.

Burchert, F., Meißner, N., Holzinger, P., and De Bleser, R., 2007. What constrains sentence production in agrammatism. *Journal Brain and Language*, 103(1-2), pp.80-81.

Chomsky, N., 1957. *Syntactic Structures.* Berlin: Mouton.

Chomsky, N., 1965. *Aspects of the Theory of Syntax*. Cambridge: M.I.T. Press.

Connor, L.T., Obler, L.K., Tocco., M., Fitzpatrick, P.M., Albert, M.L. 2001 Effects of Socioeconomic Status on Aphasia Severity and Recovery. *Journal of Brain and Language*, 78(2), pp.254-257.

Fernandez, M.G., Davis., C., Molitoris. J. J., Newhart. M., Leigh. R., and Hillis. A. E., 2011. Formal Education, Socioeconomic Status, and the Severity of Aphasia After Stroke. *Archives of Physical Medicine and Rehabilitation,* 92(11), pp.1809-1813.

Fix, S., and Thompson, C.K., 2006. Morphophonological structure and agrammatic regular/irregular past-tense production. *Journal of Brain and Language*, 99(1-2), pp.155-6.

Goodglass, H., and Kaplan, E. 1983. *The Assessment of Aphasia and Other Neurological Disorders.* Baltimore, MD: Williams and Wilkins.

Isserlin, M., 1922. Ueber Agrammatismus. *Zeitschrift fur die gesamte Neurologie und Psychiatrie,* pp.628-807.

Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., Ringelstein, E.-B., and Henningsen, H., 2000. Handedness and hemispheric language dominance in healthy humans. *Brain − A Journal of Neurology,* 123(12), pp.2512-2518.

Kolk, H., & Heeschen, C. 1990. Adaptation symptoms and impairment symptoms in Broca's aphasia. *Aphasiology*, 4(3), pp.221-231.

Laser, RM., Speizer, AE., Festa., JR., Krakauer. JW., and Marshall., RS., 2008. Variability in language recovery after first-time stroke. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(5), pp.530-534.

Lexical Analysis Software Ltd*., Scott, M., 2012. WordSmith Tools Manual software for finding word patterns.* [online] Available at:
<http://www.lexically.net/downloads/version6/HTML/?wshell.htm> [Accessed September 2012].

McCaffrey, P., 1998. Chapter 7. Specific Syndromes: The Nonfluent Aphasias, *CMSD 636 Neuropathologies of Language and Cognition.* [online] California State University, Chico Available at: <http://www.csuchico.edu/~pmccaffrey/syllabi/SPPA336/336unit7.html> [Accessed December 2012].

MacWhinney, B., Fromm, D., Holland, A., Forbes, M. and Wright, H., 2010. Automated analysis of the Cinderella story. *Aphasiology*, 24 (6-8), pp.856-68.

MacWhinney, B., Fromm, D., Holland, A. and Forbes, M., 2011. Aphasia Bank: Methods for studying discourse. *Aphasiology*, 25 (11), pp.1286-1307.

Menn, L., Obler, L. K. and Miceli, G., 1990. *Agrammatic Aphasia -A cross-language narrative sourcebook 1.* Amsterdam: John Benjamin Publishing Company.

National Stroke Association, 2012. *Recovery After Stroke: Thinking and Cognition.* [online] Available at:
<http://www.stroke.org/site/DocServer/NSAFactSheet_Cognition.pdf?docID=986>
[Accessed December 2012].

Oxford University Press, 2011. *Definition of metathesis noun from the Oxford Advanced Learner's Dictionary*. [online] Available at:
<http://oald8.oxfordlearnersdictionaries.com/dictionary/Metathesis> [Accessed June 2012].

Pedersen, P.M., Vinter, K., & Olsen, T.S., 2004. Aphasia after stroke: Type, severity and prognosis. The Copenhagen aphasia study. *Cerebrovascular Disease*, 17, pp.35-43.

Saffran, E., Schwartz, M., & Marin, O., 1980. The word order problem in agrammatism. I. Comprehension. *Journal of Brain and Language*, 10, pp.263-280.

Soares, E.C.S., Ortiz, K.Z., 2008. Influence of brain lesion and educational background on language tests in aphasic subjects, *Journal of Dementia and Neuropsychologia,* 2(4), pp.321-327.

Steele, R.D., 2007. *Improvements in Chronic Broca's Aphasia at Impairment and Functional Communication Levels following SGD Practice and Use*. Paper presented at the 2007 Clinical AAC Research Conference, Lexington, KY. 2007. [pdf] Available at: <http://www.aacinstitute.org/CAAC/2007/AcceptedSubmissions/Steele.pdf> [Accessed December 2012].

Tainturier, M. J., Tremblay, M., and Lecours, A, R., 1992. Educational Level and the Word Frequency Effect: A Lexical Decision Investigation, *Journal of Brain and Language*, 43(4), pp.460-474.

Talkbank, 2012a. *Talkbank.* [online] Carnegie Mellon University, the University of Pennsylvania and Stanford. Available at: <www.talkbank.org> [Accessed May 2012].

Talkbank, 2012b. *Aphasiabank / Talkbank.* [online] Carnegie Mellon University, the University of Pennsylvania and Stanford. Available at: <http://talkbank.org/AphasiaBank/> [Accessed May 2012].

Talkbank, 2012c. *CHAT Error Coding for Aphasiabank.* [online] Carnegie Mellon University, the University of Pennsylvania and Stanford. Available at: <http://talkbank.org/AphasiaBank/errors.doc> [Accessed May 2012].

Talkbank, 2012d. *APHASIABANK PROTOCOL DESCRIPTION.* [online] Carnegie Mellon University, the University of Pennsylvania and Stanford. Available at: <http://talkbank.org/AphasiaBank/protocol/description.doc> [Accessed May 2012].

Talkbank, 2012e. *Protocol Pictures | Aphasiabank | Talkbank.* [online] Carnegie Mellon University, the University of Pennsylvania and Stanford. Available at: <http://talkbank.org/AphasiaBank/protocol/pictures.html> [Accessed May 2012].

Talkbank, 2012f. *AphasiaBank/English/Aphasia/ | Talkbank Transcript Browser.* [online] Carnegie Mellon University, the University of Pennsylvania and Stanford. Available at: <http://talkbank.org/browser/index.php?url=AphasiaBank/English/Aphasia/> [Accessed May 2012].

Talkbank, 2012g. *Adler04a.cha.* [online] Carnegie Mellon University, the University of Pennsylvania and Stanford. Available at: <http://talkbank.org/browser/index.php?url=AphasiaBank/English/Aphasia/Adler/adler04a.cha> [Accessed May 2012].

Talkbank, 2012h. *Scale05a.cha.* [online] Carnegie Mellon University, the University of Pennsylvania and Stanford. Available at: <http://talkbank.org/browser/index.php?url=AphasiaBank/English/Aphasia/SCALE/scale05a.cha>[Accessed May 2012].

Whitaker, H. A. 1971. *On the representation of language in the human brain*. Edmonton: Linguistic Research.

Wikipedia, 2012. *Western Aphasia Battery.* [online] Available at: <http://en.wikipedia.org/wiki/Western_Aphasia_Battery> [Accessed December 2012].

Williamson, G., 2009. *Type-Token Ratio (TTR).* [pdf] Available at: <http://www.speech-therapy-information-and-resources.com/downloads/type-token-ratio.pdf> [Accessed September 2012].

www.Your-Neurologist.com, 2012. *Left Handedness.* [online] Available at: <http://www.your-neurologist.com/left-handedness.html> [Accessed May 2012].