THE EFFECT OF VARYING CHANNEL CAPACITIES ON ARTIFICIAL GRAMMAR LEARNING

A comparison between people with and without dyslexia

In their language development, children are able to extract rules from the limited input they hear extremely fast. They are able to apply those rules to new strings of words, which they never heard before: a process known as 'the logical problem of language acquisition'. The entropy model for linguistic generalizations addresses this problem and states that the process of making generalizations is influenced by the input complexity (entropy) and the channel capacity. Whenever the channel capacity is exceeded by the input complexity, the process of making generalizations is thought to increase. In order to learn more about the point at which children start to extract rules from the input, a previous artificial grammar study is used as a model for language acquisition in children (Rădulescu, Wijnen, & Avrutin, 2014). A tendency of making more generalizations when the entropy increased and channel capacity was kept constant, is reported. The present study investigates the effect of group (dyslexics vs. non-dyslexics) on the process of making generalizations, when entropy was kept constant, in order to learn more about the effect of varying channel capacities. The low entropy condition of the original experiment was used in this follow-up study, as dyslexics were expected to make more generalizations than non-dyslexics when confronted with a low amount of entropy. This is due to the assumption that participants with dyslexia are hypothesized to have smaller channel capacities than participants without dyslexia, due to either weaker working memories or problems in procedural learning. Dyslexics' channel capacities were thus expected to be exceeded by the input and dyslexics were expected to abstract more generalized rules from the input as a consequence. Results indicated a tendency in dyslexics of making more generalizations than non-dyslexics, but no significant differences between groups were found. This is probably due to power issues of the results, as the sample sizes in the present study were relatively small. Therefore, people with dyslexia are still hypothesized to have smaller channel capacities than people without dyslexia.

MANOUK VAN KERKHOF, 4175530

SUPERVISORS: SILVIA RĂDULESCU & FRANK WIJNEN

Bachelor thesis, 2015-2016, Utrecht University

Table of contents

1. Introduction

In order to become native speakers of a language, one of the things children have to learn is a complex set of rules (grammar). Children are able to master these rules very easily in a short period of time. Considering the relatively limited input children hear from their environment and the ease with which they learn the language rules, there seems to be a problem in explaining how they do so. Children seem to effortlessly extract rules from the limited input they get: a pattern found in children's language development from all over the world. This problem is known as 'the logical problem of language acquisition' (Fromkin, et al., 2000) and this problem is still unsolved, despite many years of research.

Although very much of the language development puzzle is yet to be solved, some things are already known. At some point in their language development, children have to extract generalized rules from the input they hear. This can be seen for example when children learn the past tense of verbs. For example, many children acquiring English follow the pattern of using the correct irregular past tense at first, for instance *went*, and using the wrong form *goed* afterwards. Such errors are called overregularizations (Hoff, 2014). Overregularizations show that children do not solely rely on remembering every single word (or sound) they hear, because they do not hear those wrong forms in their input and thus cannot remember them. The fact that children use the correct past tense prior to the wrong past forms furthermore shows that they do know the correct past form at first (Hoff, 2014). At some point children move from memorizing specific words to extracting generalized rules from the input, after which they will make some overregularization errors. After a while, children will learn the exceptions to the generalized rules they have learned.

Although we do know that at some point children start to extract rules from the input they hear and stop relying on remembering all the specific words they hear alone, we don't know yet at which point this rule extracting exactly starts. In order to find out some more about what triggers rule extraction in children's language acquisition, Rădulescu, Wijnen & Avrutin (2014) developed a theory on linguistic generalizations based on entropy and channel capacity. In their research they tried to find out what triggers and limits the inductive leap from memorizing specific items to extracting general rules, according to an artificial grammar experiment in adults, which is used as a model for language acquisition in children. The authors state that the inductive leap is influenced by the complexity of the input and the processing limitations of the human brain. In order to determine the complexity of the input they use entropy as a measure, which is an information-theoretic concept. In order to understand the statements of Rădulescu et al., it is important to understand those information-theoretic concepts in more detail.

2. Information theory

Information theory tries to explain how information is processed through different 'machines' (e.g. the human brain). First, the concept of 'information' should be defined. In information theory, information is seen as something you did not know before: information is a surprise. For example, if you have a look at your watch when you already know what time it is, this gives you no information, because your watch does not tell you something you did not know before. In contrast, if you do not know what time it is and have a look at your watch, it does give you information. Information exchange only occurs when there is an uncertainty, and the amount of exchanged information is determined by that amount of uncertainty. Information can be seen as the reduction of uncertainty (Van Ewijk, 2013). This uncertainty can be measured in logarithms, and is defined by:

$$U = -\log_2 k$$

where U is the measure of uncertainty and k is the number of possible outcomes. Uncertainty is measured in bits, which is a contraction of *binary* and *digit*. Bits have a value of either 0 or 1 and are combined into possible configurations to encode possible outcomes of information (Van Ewijk, 2013).

In order to understand to what extent information could be transferred error-free through machines, Shannon (1948) created the concept of entropy. According to Shannon, it is not the transmission rate of information that predicts the amount of errors in the transmission, as was stated earlier by Nyquist and Hartley (cited in Van Ewijk, 2013). Shannon suggests that the complexity of the information predicts the amount of errors in the transmission of information. Shannon named this complexity level 'entropy' (H) and created a model to calculate the entropy level. This entropy level is based on a statistical measure for probability and states that every message has a certain probability of occurrence, or an uncertainty of occurrence, depending on the entropy level (Van Ewijk, 2013). Probabilities are numbers between 0 and 1. A probability of 0 shows that it is improbable for a value to occur, while a probability of 1 shows that a value will certainly occur. The formula suggested to measure the uncertainty (U) assumes that all outcomes are equally likely to occur, but for most situations this is not the case. Shannon (1948) therefore created a formula to calculate the entropy of an information source, which assumes that not all outcomes are equally likely to occur. In order to determine the uncertainty of a whole set, it is necessary to determine the uncertainty of the individual elements of that set, as some elements could be more likely to occur than others. When the uncertainties of all of the individual elements of a set are determined, the entropy of the whole set can be determined. Entropy is a measure for the average uncertainty, or average information, of the input (Shannon, 1948). Entropy can be calculated as follows:

$$H(x) = -\sum_{i=1}^{n} p(x_i) \log p(x_1)$$

where *H* is a measure of the entropy, *p(x)* is the probability of occurrence of any value of *x*, and *log* should be read as *log* to the base 2.

Apart from entropy, Shannon's theory (1948) is also based on 'channel capacity'. Aside from calculating how much information is being transmitted, it is important to know how much information a machine is able to process per unit of time. Channel capacity is a measure for the maximum amount of entropy a machine can process error-free, and can be determined by presenting varying amounts of entropy to the machine. As long as the entropy of the machine's outcome increases linearly with the entropy of the input, the machine's channel capacity is not exceeded (Van Ewijk, 2013). Whenever the channel capacity of a machine is exceeded by the entropy of the input, the amount of entropy has to be reduced in order to process the information error-free. This can be done by regularizing the information (i.e. reorganizing the information based on regularities in it), so that the entropy of the input is reduced and does not exceed the machine's channel capacity any longer (Rădulescu, Wijnen, & Avrutin, 2014). In order to reduce the entropy, information can be regularized by extracting rules, for example. Rădulescu et al. tried to find out at what point rule extracting starts to increase in the human artificial grammar learning process. In order to do this, the authors created an entropy model for linguistic generalizations.

3. An entropy model for linguistic generalizations

Rădulescu et al. combine the entropy of the input with channel capacity in their model for linguistic generalizations and state that "generalization is a cognitive mechanism that results from the interaction of input complexity (entropy) and the processing limitations of the human brain, i.e. limited channel capacity" (Rădulescu, Wijnen, & Avrutin, 2014, p. 3). This model applies to language acquisition because of the fact that in order to become fluent speakers of their mother tongue, children have to memorize specific linguistic items (e.g. English – *dog, house, table*) and have to acquire linguistic categories (e.g. the category of nouns). Children cannot remember every single noun and its specific characteristics, because eventually their channel capacity will be exceeded. In order to reduce the entropy of their input, children eventually have to extract general characteristics from the input for the category of nouns. To become fluent speakers of their mother tongue, this has to be done for every linguistic category. Although it is known that this process has to take place at some point in a child's language development, it is unknown what exactly triggers the rule extracting. That is what Rădulescu et al. tried to find out via an artificial language experiment with adults. We assume that adult artificial grammar learning is a representative model for language acquisition in children.

In order to determine at what point humans move from memorizing specific items to extracting general rules, Rădulescu et al. (2014) make a distinction between two types of abstractions: pattern-based abstractions and category-based abstractions. A relation between perceptual characteristics of elements, for example a relation based on physical identity (e.g. *ba-ba*, in which *ba* is followed by *ba*), is a pattern-based abstraction. On the other hand, a relation between abstract variables (e.g. *A-A-B* abstracted from *babalu*, *kokofe, dedemo*, etc.) is a category-based abstraction. Pattern-based abstractions correspond to memorizing specific items out of the input and can only be applied to the specific items they are related to. In contrast, category-based abstractions correspond to extracting general rules from the input and can be applied to new strings of words. In order to determine at what point humans move from memorizing specific items to extracting general rules from the input, the authors attempted to discover what triggers and what limits the process of making pattern-based abstractions towards making category-based abstractions.

In their experiment, Rădulescu et al. (2014) used an artificial language, which they created. This miniature artificial grammar consisted of words with an XXY structure, where each letter represents a set of syllables (e.g. *daadaalie*, *puupuuvee*, *keekeemuu*). In the experiment, participants were exposed to grammatical stimuli (i.e. following an XXY structure) in the training phase at first. Subsequently participants had to complete a test in which both grammatical and ungrammatical (i.e. following an XYZ structure; e.g. *hiedaareu*) stimuli were presented, which consisted of both trained (i.e. occurred in the training phase) and untrained (i.e. novel) syllables. Grammatical test items with novel syllables were most useful to test category-based abstractions. Participants had to judge whether the items were correct according to the artificial language they were exposed to in the training phase. This process of training and test phases was repeated three times, after which a final test was completed.

In order to do determine at what point humans start to make category-based abstractions, Rădulescu et al. varied the amount of entropy in the input and hypothesized to keep the channel capacity constant. We assume that the processing capacity of the human mind matures in time, and so does the channel capacity. By using participants of roughly the same age, the authors kept the channel capacity of the participants as constant as possible. By varying the amount of entropy in the input, they tried to find out what its effect is on making pattern-based or category-based abstractions while the channel capacity was kept constant. In order to vary the amount of entropy over the different entropy conditions in their experiment, the authors varied the number of syllables that occurred in the artificial grammar over the conditions, where a higher number of syllables corresponded with a higher entropy. The results of the experiment showed that participants tended to make more category-based perceptions (i.e. accepted grammatical stimuli with both trained and untrained syllables) when the entropy of the input was higher, and tended to make more pattern-based abstractions (i.e. mainly accepted grammatical stimuli with trained syllables) with a lower entropy in the input (Rădulescu, Wijnen, & Avrutin, 2014).

In this experiment, participants were hypothesized to have roughly the same channel capacity due to the fact that they were of the same age. It is interesting though to have a look at varying channel capacities and their effect on the point at which humans start to abstract generalized rules. The expectation would be that participants with a lower channel capacity sooner tend to extract generalized rules from the input than the participants in the original experiment, because of the fact that the input will exceed their channel capacity earlier. In order to test this, mainly the low entropy condition of the original experiment will learn us more about the point at which humans with smaller channel capacities start to make generalizations, as participants in the original experiment were shown to abstract rules in the low entropy condition to a lesser extent than in the medium and high entropy conditions. When humans with smaller channel capacities than in the original experiment will participate in the low entropy condition, this will prove whether these participants indeed abstract more rules from the low entropy input than participants in the original experiment. This can be investigated by testing participants who are considerably younger (Rădulescu, Wijnen, & Avrutin, 2014), on the assumption that channel capacity increases with age.

Finding out what the effect of varying channel capacity is on the process of extracting generalized rules from the input, could also be done by testing dyslexic participants. Next to the well-known reading and writing difficulties, people with dyslexia are also known to have problems with recognizing patterns and rules from strings of input. The underlying learning problems might explain reading and writing difficulties that come with dyslexia. The exact causes of dyslexia are unknown and many theories have been proposed over the years. According to the *Procedural deficit hypothesis*, the problems are due to a limited procedural learning system caused by abnormal brain structures (Nicolson & Fawcett, 2007). This causes people with dyslexia to have difficulties in automatizing observed patterns. In order to become a competent reader and writer, it is important to find regularities in a language's orthography. Subsequently, these patterns should be memorized in order to automatize the process of recognizing and producing these patterns in written language. This process does not seem to work properly in people with dyslexia. Because of the problems people with dyslexia have in automatizing observed patterns, they have, among others, weaker reading and writing skills than people without dyslexia according to the Procedural deficit hypothesis. The problem in recognizing patterns and rules could also be due to a problem in the working memory of dyslexic people, which seems to be weaker than in non-dyslexic people (Schuchardt, Bockmann, Bornemann, & Maehler, 2013). Due to weaker working memory, people with dyslexia would not be able to memorize input long enough to find regularities in it.

Anyway, whatever the precise underlying cause may be for the problems associated with dyslexia, the problems in procedural learning and working memory can be construed as limitations of the channel capacity. Based on the observation that the procedural learning system and working memory are weaker in dyslexics than non-dyslexics, it makes sense to hypothesize a smaller channel capacity in dyslexics. Because of the idea that dyslexics have a smaller channel capacity than (adult) non-dyslexics, dyslexics are thought to have a more 'child-like' learning pattern: Their channel capacity is hypothesized to be exceeded sooner than non-dyslexics' channel capacity and this may cause category-based abstractions to arise earlier (i.e. when confronted with a relatively low amount of entropy in the input) than found in the participants of the original experiment (Rădulescu, Wijnen, & Avrutin, 2014). Finding out what kind of learning pattern dyslexic participants show compared to non-dyslexic participants, when confronted with a low amount of entropy in the input, is the goal of this thesis.

4. Research question and hypotheses

The research question of this thesis is as follows: is there a difference between people with and without dyslexia regarding rule (or pattern) abstraction, when confronted with an artificial language with a low entropy? This question leads to the following hypotheses:

(a) Participants with dyslexia are expected to extract rules from the input, due to exceeded channel capacities. This will cause them to reorganize the information of the input and to find regularities in it.

(b) Participants without dyslexia are expected to perform as in the previous experiment (Rădulescu, Wijnen, & Avrutin, 2014; 2015), i.e. they are not expected to extract rules from the input, due to non-exceeded channel capacities.

In order to test these hypotheses, four different types of test stimuli will be used. Grammatical stimuli (i.e. following an XXY structure) with both trained and untrained syllables and ungrammatical stimuli (i.e. following an XYZ structure) with both trained and untrained syllables will be presented to the participants in the test phases. Especially grammatical stimuli with novel syllables will be useful to test whether participants abstracted generalized rules from the input, due to the fact that participants cannot judge them to be correct based on memorization as a consequence of exposure in the training phases. According to the hypotheses, participants with dyslexia are expected to perform better than participants without dyslexia on the grammatical test items with untrained syllables, due to the assumption that their channel capacity will be exceeded by the input, and they will have to reorganize the information and find regularities in it as a consequence. Participants without dyslexia are not expected to reorganize the information of the input, due to the assumption that their channel capacity will not be overloaded by the input. As a consequence, participants without dyslexia are not expected to extract generalized rules from the input and are expected to perform weaker than participants with dyslexia on XXY test items with novel syllables.

5. Method

5.1 Participants

Both dyslexic and non-dyslexic participants were recruited via various posts on social media and the author's social network. A total of 23 Dutch speaking adults (age range 19-25), who were not familiar with the research topic, participated in the experiment. Dyslexics and non-dyslexics were matched on gender, age and educational background. Both dyslexic (n=12) and non-dyslexic (n=11) subjects were accepted in the experiment, as long as they had no known hearing impairment or attention deficit. Subjects participated in the experiment for free.

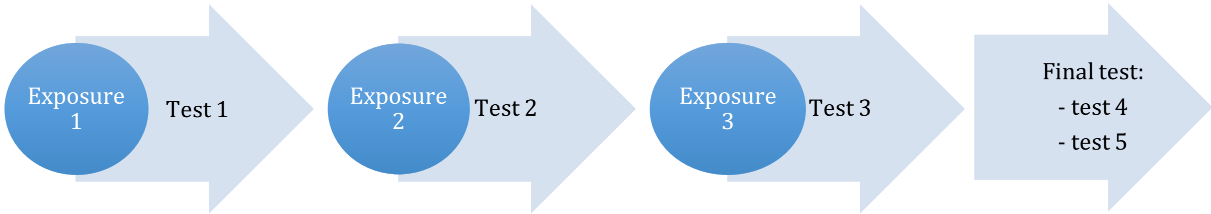
Statistical tests were carried out in order to check whether the groups were comparable with regard to age, gender, and educational background. An alpha level of *.05* was used for these tests. An Independent-Samples T Test showed no significant difference in age between dyslexic (M = 22.50, SD = 2.35) and non-dyslexic (M = 22.18, SD = 1.78) participants (t = 0.38; df = 20.30; p = .717). Fisher's exact Test showed no significant difference (p = .611) in gender between dyslexics (female/male = 8/4) and non-dyslexics (female/male = 7/4). Furthermore, Fisher's exact Test showed no significant difference (p = .999) in education level between dyslexics (MBO/HBO/WO = 1/6/5) and non-dyslexics (MBO/HBO/WO = 1/6/4). The groups are therefore assumed to be similar, and differences in test results thus cannot be explained by between-group differences in age, gender or educational background.

5.2 Training stimuli

The same stimuli were used as in the low entropy condition in the experiment of Rădulescu et al. (2015). This means that the set of training stimuli has an entropy of 2.8 bits (see Appendix 1 for the complete entropy calculations). All syllables were recorded one by one in a sound-proof booth by a female Dutch native speaker. She was instructed to use the same intonation for each syllable. The set of training stimuli consists of words with three syllables, which all start with a consonant that is followed by a long vowel (e.g. *daa*, *teu*) in order to resemble common Dutch syllable structure. Syllables were checked on actual existence in Dutch and were excluded from the artificial language if they did occur in Dutch words. All stimuli follow an XXY structure, where each letter represents a set of syllables. A subset of the syllables in the artificial language is used in the X-positions, while another subset is used in the Y-positions. Therefore, X-syllables never occur in an Y-position, and vice versa. Examples of stimuli are *daadaalie* and *teuteureu*. The complete set of training stimuli consists of 28 strings, with 7 X-syllables and 7 Y-syllables (see Appendix 2 for the complete list of stimuli). Stimuli are presented in a randomized order per participant. After all training phases, participants will have been exposed to 84 training stimuli in total.

5.3 Procedure

Subjects participated in the experiment at home or at school, with as few distracting factors in the surroundings as possible. In order to classify participants as dyslexic and to make a comparison between the results of people with and without dyslexia on the artificial grammar task, the experiment also consisted of two reading tasks, a verbal competence task and a memory task. During the experiment subjects participated in the artificial grammar task at first. In this task the participants had to listen to 28 stimuli in all 3 training phases. These stimuli were words which were correct according to the artificial grammar that was constructed by Rădulescu et al. (2014). After a training phase, the test phase started. In this phase the participants had to decide whether the 4 given test items could occur in the artificial grammar they were exposed to in the training phase. This procedure was repeated three times, after which there was a final test. In this final test participants were confronted with another set of 8 test items (consisting of two separate tests, to be called 'test 4' and 'test 5'), about which they had to decide whether they were correct according to the artificial grammar. Intermediate test phases were included in the design to be able to have a closer look at the learning curve and to check whether participants tend to generalize more when exposure is extended. The procedure of the artificial grammar task is shown in the timeline below:



Subsequently participants had to complete a *visual forward digit span* task to test their working memory. In this task participants had to reproduce a series of digits that were shown to them on a computer screen. The shown set of digits was extended with one digit per test phase. This process continued for as long as the participant did not make a mistake or, alternatively, up to a digit span of ten digits. This task is especially interesting to compare the results of the dyslexic and non-dyslexic participants, because working memory is assumed to be related to channel capacity. A difference in the outcomes of this visual forward digit span task between groups might explain (part of) the possible differences in the outcomes of the artificial grammar task. Dyslexics are expected to show significant poorer performance than non-dyslexics, based on earlier findings that dyslexics have a weaker working memory than non-dyslexics (Schuchardt, Bockmann, Bornemann, & Maehler, 2013).

Finally, the participants were tested for their reading abilities and verbal competence. In order to do this, they had to complete the 'Een Minuut Test' (EMT; (Brus & Voeten, 1973) at first. This is a list of existing words, which participants had to read out loud as fast and accurate as they were able to within one minute. After this, the participants' verbal competence was tested with the verbal competence task of the WAIS IQ-test (Wechsler, 1955). Participants had to name similarities between pairs of two words in this test, e.g. '*car* vs. *airplane*' and '*day* vs. *night*'. Finally, participants had to complete the Klepel (Van den Bos, Lutje Spelberg, Scheepstra, & De Vries, 1994). This is a list of phonotactically legal non-words, which participants had to read out loud as fast and accurate as they were able to within two minutes. These tests were used to classify participants as dyslexic. This was done according to the norms of Kuijpers, Van der Leij, Van Leeuwen, Ter Keurs, Schreuder & Van den Bos (2003), who set norms for classifying dyslexia for research purposes with the named tests. Determining whether participants are dyslexic is necessary to compare the artificial grammar task and visual forward digit span task results for the dyslexic and non-dyslexic groups. Dyslexics are expected to perform significantly more poorly than non-dyslexics on both the EMT and Klepel. No significant differences are expected in performance between groups for verbal competence, because participants were matched on educational background.

5.4 Test stimuli and performance predictions

The set of test stimuli consists of four different kinds of stimuli: grammatical stimuli with trained syllables (i.e. syllables that occurred in the training set), grammatical stimuli with untrained (i.e. novel) syllables, ungrammatical stimuli with trained syllables and ungrammatical stimuli with untrained syllables. Grammatical test stimuli follow an XXY structure (e.g. *daadaalie*, *teuteureu*), while ungrammatical test stimuli follow an XYZ structure (i.e. comprising 3 different syllables; e.g. *hiedaareu*, *keefoovee*). The complete list of test stimuli comprised 20 items (see Appendix 3). Each of the four types of test stimuli is designed to test a specific mechanism of rule extraction (Rădulescu, Wijnen, & Avrutin, 2014). According to the entropy model for linguistic generalizations, each type of test stimuli should show a particular type of learning tendency, which will be discussed below. These performance predictions are based on predictions made by Rădulescu et al. according to their different entropy conditions.

Type_1 comprises XXY stimuli with both trained X and trained Y syllables, which should be considered correct. This type of test items is supposed to check learning of the trained strings and structure and is a positive test case, because participants have heard the XXY structure as well as the syllables. As a consequence, participants could judge type_1 stimuli to be correct, based on memorization from the training phase input alone. For this type of test stimuli, it is thus not necessary to have abstracted generalized rules from the input in order to give the correct answer. Both the dyslexic and non-dyslexic participants are therefore expected to show a high performance for type_1 stimuli. Dyslexics are expected to show high performance based on strongly or at least satisfactorily developed category-based abstractions, due to an exceeded channel capacity and rule abstraction. Non-dyslexics are expected to show high performance based on strongly developed pattern-based abstractions, due to non-exceeded channel capacity and no generalized rule abstraction. Consequently, both groups are expected to perform significantly above chance level, with no between-group differences.

Type_2 consists of XYZ stimuli with both untrained X and untrained Y syllables, which should be judged as *in*correct. This type of test stimuli is also designed to test learning of the trained strings and structure, but is a negative test case complementary to type_1. If the hypothesized tendencies for type_1 are correct, type_2 results should therefore be consistent with the results of the type_1 stimuli. Participants cannot consider type_2 items correct based on either learned syllables or structure, because they did not hear the structure or syllables in the training phases. Both the dyslexic and non-dyslexic group are expected to show high performance for type_2 stimuli. Dyslexics probably consider type_2 items incorrect based on strongly or at least satisfactorily developed category-based abstractions. Non-dyslexics probably consider type_2 stimuli incorrect based on strongly developed pattern-based abstractions. As a result, both groups are expected to perform significantly above chance level, with no between-group differences.

Type_3 consists of XXY stimuli with both untrained X and untrained Y syllables, which should be judged as correct. Type_3 is a positive test case and is designed to check whether participants extracted a generalized rule from the training stimuli. This type of test items is most useful to test rule abstraction and thus the difference between participants with and without dyslexia regarding generalizations, due to the fact that participants should consider type_3 items correct based on the learned structure alone, as participants have not heard the syllables in the training phases. Both groups are expected to perform significantly above chance level, based on Rădulescu et al.'s findings (2015), but with between-group differences. The highest number of correct answers is expected to be given by the dyslexics, due to satisfactorily developed category-based abstractions. The lowest number of correct answers is expected to be given by the non-dyslexics, due to strongly developed pattern-based and weakly developed category-based abstractions.

Type_4 comprises XYZ stimuli with both trained X and trained Y syllables, which should be considered *in*correct. Type_4 is designed as a negative test case complementary to type_3. If the hypothesized tendencies for type_3 stimuli are correct, type_4 results should be consistent with the results of type_3 stimuli. Participants should consider type_4 items incorrect, based on the learned structure. The pattern-based and category-based abstractions should work against each other, because the trained syllables tend to drive pattern-based abstractions to accept the stimuli, while the structure drives category-based abstractions to reject the stimuli. As a result, difference in performance level is expected between groups. Dyslexics are expected to give the highest numbers of correct answers, due to a higher tendency to make category-based abstractions than non-dyslexics. As a result, dyslexics will have abstracted the grammatical XXY pattern from the input, and their memory traces of the trained syllables will be weaker than that of the non-dyslexics. Non-dyslexics are expected to show lower performance than dyslexics, due to strongly developed pattern-based abstractions, and thus stronger memory traces of the trained syllables than that of non-dyslexics. Consequently, they are expected to show weaker performance than participants with dyslexia. Both groups are expected to perform above chance-level, according to the findings of Rădulescu et al. (2015)

6. Results

6.1 Reading tests and digit span

The percentile scores for the EMT, Klepel and the verbal competence task were computed according to the total scores. Participants were regarded as dyslexic when they met at least one of the following criteria:

1. Percentile score of 10 for EMT and/or Klepel;
2. Percentile score of 20 for both EMT and Klepel;
3. Difference of 60 in percentile scores for EMT and/or Klepel in comparison with verbal competence score.
   (Kuijpers, et al., 2003)

All of the dyslexic participants met the first two criteria: they either had a percentile score of 10 for the EMT and/or Klepel, or had a percentile score of 20 for both the EMT and the Klepel. None of the participants were classified as dyslexic based on a difference of 60 in percentile scores for EMT and/or Klepel in comparison with the verbal competence task score alone. After this analysis, two out of thirteen self-claimed dyslexic participants were assigned to the non-dyslexic group. Vice versa, one out of twelve self-claimed non-dyslexic participants was regarded as a dyslexic participant in this study.

Subsequently, both groups were compared on EMT, Klepel, verbal competence and digit span scores. Table 1 shows mean scores and standard deviations for both groups and the outcomes of statistical tests. An alpha level of *.05* was used for all statistical tests. Independent-Samples T tests showed significant differences between groups for EMT ($p < .001$) and Klepel ($p < .001$) scores, but no significant difference between groups for verbal competence ($p = .629$) and digit span ($p = .164$) scores. EMT, Klepel and verbal competence scores are as expected, with significant differences between groups for the EMT and Klepel, but not for verbal competence scores. Against the odds, digit span scores do not differ significantly between groups however.

Table 1

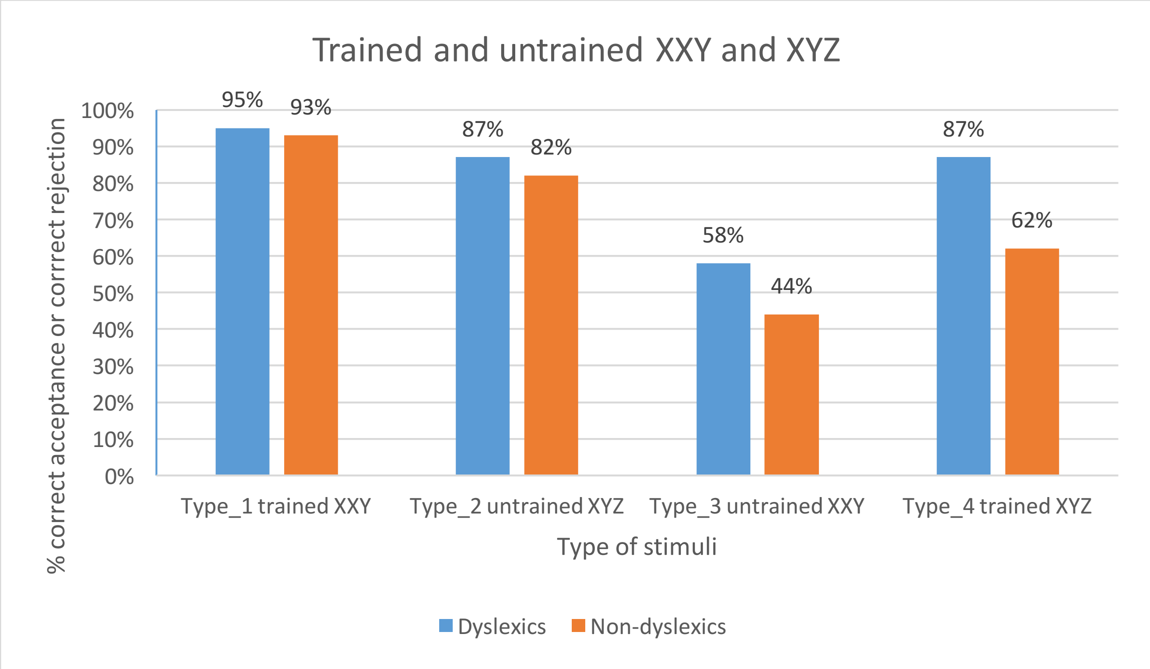
Mean Scores (and Standard Deviations) for EMT, Klepel, Verbal Competence and Digit Span for Dyslexics and Non-dyslexics. Additionally, Outcomes of Statistical Analyses are given.

Test	Dyslexics (N = 12)	Non-dyslexics (N = 11)	t	df	p
EMT	78.67 (9.87)	104.00 (10.74)	5.89	21	< .001
Klepel	62.83 (16.22)	93.09 (14.03)	4.76	21	< .001
Verbal Competence	19.92 (3.99)	19.18 (3.09)	0.49	21	.629
Digit Span	7.00 (1.41)	8.00 (1.89)	1.44	21	.164

6.2 Artificial grammar learning

In order to test the effect of group (participants with dyslexia versus participants without dyslexia) on the process of making generalizations, when input complexity was kept constant, performance levels of both groups for the different test stimuli types were compared. An alpha level of *.05* was used for all statistical tests. The performance levels (in percentages of correct answers) of both groups for each test stimuli type are presented below. Figure 1 shows the mean performance for both groups for all types of test stimuli in percentages of correct answers (i.e. correct acceptance for type_1 and type_3, and correct rejection for type_2 and type_4).

Figure 1
Percentages of Correct Acceptance for Type_1 and Type_3 and Percentages of Correct Rejections for Type_2 and Type_4 for both groups.



For type_1stimuli (XXY with trained syllables), dyslexics showed a mean performance level of 95% (M = 4.75, SD = 0.45). Non-dyslexics showed a mean performance of 93% (M = 4.64, SD = 1.21). An Independent-Samples T Test showed no significant difference between performance levels of both groups (t = 0.30; df = 21; p = .764). One-Sample T Tests indicated a significant above-chance performance for both dyslexics (t = 36.38; df = 11; p < .001) and non-dyslexics (t = 5.88; df = 10; p < .001).

Type_2 stimuli (XYZ with untrained syllables) yielded mean performance of 87% for dyslexics (M = 4.33, SD = 0.65) and 82% for non-dyslexics (M = 4.09, SD = 0.94). An Independent-Samples T Test showed no significant difference between performance levels of both groups (t = 0.72; df = 21; p = .478). A significant above-chance level was indicated by One-Sample T Tests for dyslexics (t = 9.75; df = 11; p < .001) and non-dyslexics (t = 5.59; df = 10; p < .001).

For type_3 items (XXY with untrained syllables), which show best whether participants abstracted generalized rules from the input, dyslexics showed a mean performance level of 58% (M = 2.92, SD = 1.83). Non-dyslexics showed a mean performance level of 44% (M = 2.18, SD = 2.04). Although mean performance levels per group seem to differ considerably, an Independent-Samples T Test indicated no significant difference in mean group performances (t = 0.91; df = 21; p = .373) however. One-Sample T Tests indicated no significant above-chance performance for dyslexics (t = 0.79; df = 11; p = .447) and no significant below-chance performance for non-dyslexics (t = 0.52; df = 10; p = .616).

Type_4 stimuli (XYZ with trained syllables) evoked mean performance level of 87% for dyslexics (M = 4.33, SD = 0.49) and 62% for non-dyslexics (M = 3.09, SD = 1.92).

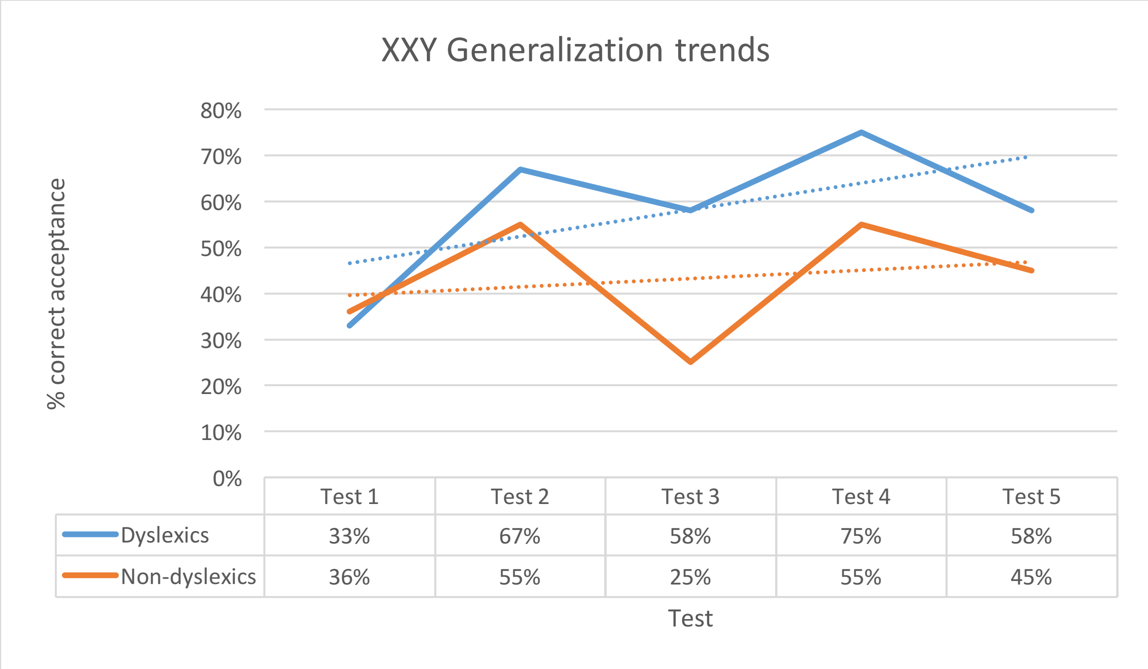
An Independent-Samples T Test showed no significant difference between mean performance levels of both groups ($t = 2.08$; $df = 11.20$; $p = .061$). A significant above-chance performance was indicated by One-Sample T Tests for dyslexics ($t = 12.90$; $df = 11$; $p < .001$), but not for non-dyslexics ($t = 1.02$; $df = 10$; $p = .332$).

In order to test whether one group showed higher variability between performance levels for type_1 (XXY with trained syllables) and type_3 stimuli (XXY with untrained syllables) than the other group, mean differences between performance levels for both groups were compared. High variability between performance levels for these stimuli indicates a low level of rule abstraction, as rule abstraction would lead to equal performance levels for both types of test items, because of the fact that both types of stimuli would be judged correct according to abstracted rules. The mean difference between type_1 and type_3 for dyslexics ($M = 1.83$, $SD = 1.70$) was lower than for non-dyslexics ($M = 2.45$, $SD = 1.97$), but an Independent-Samples T Test indicated no significant difference for both groups in mean differences ($t = 0.81$; $df = 21$; $p = .425$).

In order to test whether one group showed higher variability between performance levels for type_1 (XXY with trained syllables) and type_2 (XYZ with untrained syllables) stimuli, mean differences between performance levels for both groups were compared. High variability between performance levels for these stimuli indicates a low level of both pattern- and category-based abstractions, as both abstractions lead to correct answers (i.e. accepting type_1 and rejecting type_2). The mean difference between type_1 and type_2 for dyslexics ($M = 0.42$, $SD = 0.51$) was lower than the mean difference for non-dyslexics ($M = 0.55$, $SD = 1.21$). An Independent-Samples T Test showed no significant effect in mean differences for both groups ($t = 0.34$; $df = 21$; $p = .740$) however.

Figure 4 shows the percentages of correct answers for type_3 (untrained XXY) stimuli detailed by intermediate tests. Results indicated a slowly increasing trend for dyslexics and an even slower increasing trend for non-dyslexics for the direction of the generalization tendency as a function of exposure.

Figure 2
Percentage of Correct Acceptance for Type_3 Untrained XXY Detailed by Intermediate Tests. The Trend Lines show the Direction of the Generalization Tendency as a Function of Exposure.



7. Discussion and conclusions

This study was a follow-up to the study of Rădulescu et al. (2014). According to the entropy model for linguistic generalizations proposed by Rădulescu et al., language development starts out with memorizing specific items out of the input, based on pattern-based abstractions. As long as the channel capacity is not exceeded by the input complexity (entropy), this process continues. Whenever the input complexity exceeds the channel capacity, pattern-based abstractions are no longer sufficient to process the input error-free. At this point, the tendency towards learning based on category-based generalizations starts to develop gradually. The authors concluded their entropy model for linguistic generalizations to be correct according to their results and suggested further research in, among others, varying channel capacities when input complexity was kept constant.

In the present study the effect of group (participants without dyslexia versus participants with dyslexia) on the process of making generalizations in an artificial grammar experiment, when input complexity was kept constant, was investigated. In order to do this, stimuli of Rădulescu et al. were used and both dyslexic and non-dyslexic participants were tested. Dyslexic participants were hypothesized to have a smaller channel capacity than non-dyslexic participants, due to either weaker working memory or problems in procedural learning. According to this hypothesis, dyslexics were predicted to show better performance than non-dyslexics on grammatical stimuli with untrained syllables, due to overloaded channel capacity which causes them to make more category-based abstractions than non-dyslexics.

Any reported differences between groups, do not necessarily indicate a difference in channel capacities between groups, however. Differences could also be due to varying competence in recognizing or memorizing auditory stimuli, for example, which may lead to varying performance levels between groups. The experiment design is not able to exclude these factors completely. However, it makes sense to hypothesize weaker competence in participants with dyslexia than participants without dyslexia in recognizing or memorizing auditory stimuli. Problems in recognizing and memorizing auditory input may also lead to problems in reading and writing, because of the lack of recognizing regularities in the input as a consequence: problems found in people with dyslexia. Therefore, it makes sense to hypothesize weaker competence in memorizing and recognizing stimuli for dyslexics. Therefore, poorer performance for dyslexics than non-dyslexics in the present study would be hypothesized, when reported differences would be caused by problems in recognizing and memorizing skills, and not by varying channel capacities. However, according to the (non-significant) higher performance of the dyslexic group, problems in recognizing or memorizing the stimuli do not seem to be likely. It seems to be plausible, therefore, that reported differences are caused by varying channel capacities between groups.

As expected, both groups showed high performance for type_1 (trained XXY) test stimuli and no significant difference was indicated between mean group performances. Furthermore, both groups showed above-chance performance for type_1 test stimuli. Considering the fact that type_2 (untrained XYZ) test items were designed as a negative test case complementary to type_1 test stimuli, and both groups showed equal performance on the two test types, performance is as was expected for both groups. Prior to the experiment, a between group difference was expected for type_3 (untrained XXY) test items. Dyslexics in fact showed higher mean performance than non-dyslexics. The analyses showed no significant between group difference in performance level however. Both groups showed performance on type_4 according to the predicted directions: dyslexics performed better than non-dyslexics. However, no significant difference was found. Altogether, the performance predictions made prior to the experiment do not seem to be completely correct. No significant differences were found in mean performance levels between dyslexics and non-dyslexics in the predicted directions. However, both groups showed performance in the predicted tendencies. The lack of significant differences might be explained by a power problem of the results, due to either the used statistical tests or the relatively small sample sizes used in this study (Dyslexic group: N = 12, Non-dyslexic group: N = 11), and relatively high variances. The hypotheses (i.e. dyslexics were hypothesized to abstract rules, and non-dyslexics were not expected to abstract rules) seem to be correct, therefore.

Moreover, the non-dyslexic group was expected to perform equal to the performance levels found in Rădulescu et al.'s experiment, as no changes were made according to those participants' channel capacity and the entropy of the input. However, type_3 (untrained XXY) stimuli evoked different performance levels in the two experiments. Rădulescu et al. reported a percentage of correct answers of 57% for type_3, while the present study remarkably found a percentage of 44% for non-dyslexic participants. Considering the fact that no changes were made in channel capacity and entropy between the two studies for non-dyslexic participants, the difference in results might be due to unfortunate test circumstances in the present study. In the original experiment participants were tested in a sound-proof booth, while in the present study participants were tested at home or at school. Although background noises were eliminated as much as possible, the test locations were not sound-proof. This may have caused the participants to be distracted, which may explain the difference in reported results. Furthermore, participants in the original experiment were reimbursed, while subjects in the present study participated for free. This may have caused participants in the original study to be more motivated than in the present experiment. In future research all participants should be reimbursed and all experiments should take place in a sound-proof booth, to rule out motivation due to reimbursement and background noise as influencing factors on the results. Non-dyslexic participants' performance levels are then expected to improve, up to the reported performance level of the original experiment. Dyslexics' performance levels are also expected to improve, as dyslexic participants may also have been distracted or suffered a lack of motivation.

The fact that dyslexics and non-dyslexics did not perform significantly different for type_3 test items, could also be explained by the fact that the two groups did not have significantly different digit span scores. In the present study we assumed that digit span scores are a proper measure for working memory, although this is not commonly agreed upon. As a consequence to this assumption, we assumed that digit span scores are related to channel capacity sizes, as working memory is related to channel capacity. Dyslexics were expected to perform significantly lower than non-dyslexics on the digit span task, due to weaker working memories reported in earlier studies (Schuchardt, Bockmann, Bornemann, & Maehler, 2013). Dyslexics were therefore expected to have smaller channel capacities. This assumption resulted in the hypothesis that dyslexics would perform significantly better on type_3 test stimuli. However, according to the fact that digit span scores did not differ significantly between groups, the hypothesized smaller channel capacity for dyslexics may have to be revised. On the one hand, non-significant differences between digit span scores may indicate that digit span is not a proper test for working memory, as dyslexics are expected to have weaker working memories than non-dyslexics. On the other hand, equal channel capacities for dyslexics and non-dyslexics explain the lack of significant difference between group performance levels for type_3 test items. But then again, sample sizes were relatively small in the present study ($n = 12$ for dyslexics; $n = 11$ for non-dyslexics). This may explain the fact that digit span scores and type_3 performance levels did not differ significantly between groups, due to power issues of the results. According to the findings of Schuchardt et al. (2013) and the findings of the present study (digit span score of 7 for dyslexics vs. 8 for non-dyslexics), digit span scores are still assumed to be a proper measure for working memory. Furthermore, digit span scores are expected to differ significantly between groups with bigger sample sizes in future research. Channel capacities thus are still hypothesized to be smaller for dyslexics than non-dyslexics.

Rădulescu et al. propose a function for the two learning mechanisms, namely pattern-based and category-based learning, which shows the effect of varying entropies on the learning mechanisms. Future research should expose both dyslexic and non-dyslexic participants to varying input complexities, in order to figure out what learning mechanism function could be made when channel capacity is smaller than in the original experiment. This function is expected to show earlier decrease of pattern-based learning and sooner increase of category-based learning for dyslexics as compared to the original function when entropy increases, due to earlier exceeded channel capacities. Findings of the present study confirm this hypothesis.

Furthermore, future research should expose both children and adults to varying amounts of entropy in order to make a function of the learning mechanisms for children. This function is also expected to show earlier decrease for pattern-based learning and sooner increase for category-based learning for children as compared to the function in the original experiment. When both functions, for children and dyslexics, are made, the functions can be compared in order to compare learning mechanisms of children to learning mechanisms of dyslexics. Considering the fact that channel capacity increases in time, (adult) dyslexics are expected to show later decrease of pattern-based abstractions and later increase of category-based abstractions than children, due to even smaller channel capacities in children.

The results of the present study do not show significantly different results between participants with and without dyslexia in abstracting generalized rules when confronted with a low entropy in the input, which is probably due to the small sample sizes in the present study. However, the fact that the results do show a trend in the predicted directions (i.e. participants with dyslexia extract rules from the input, while participants without dyslexia do not extract rules) suggests that a more extensive study, with sound-proof test locations, might show significant differences between participants with and without dyslexia according to rule abstraction.

8. References

Brus, B. T., & Voeten, M. J. (1973). Een-minuuttest. Nijmegen: Berkhout.

Fromkin, V. A., Curtiss, S., Hayes, B. P., Hyams, N., Keating, P. A., Koopman, H., . . . Szabolcsi, A. (2000). Linguistics: An introduction to linguistic theory. In V. A. Fromkin. Oxford: Blackwell Publishing Ltd.

Hoff, E. (2014). Language development. Andover: Wadsworth cengage learning.

Kuijpers, C., Van der Leij, A., Been, P., Van Leeuwen, T., Ter Keurs, M., Schreuder, R., & Van den Bos, K. P. (2003). Leesproblemen in het voortgezet onderwijs en de volwassenheid. *Pedagogische studiën, 80*(4), pp. 272-287.

Nicolson, R. I., & Fawcett, A. J. (2007). Procedural learning difficulties: reuniting the developmental disorders? *Trends in neuroscience, 30*(4), 135-141.

Rădulescu, S., Wijnen, F., & Avrutin, S. (2014). Patterns bit by bit. An entropy model for linguistic generalizations. Utrecht university.

Rădulescu, S., Wijnen, F., & Avrutin, S. (2015). Input complexity and rule induction. An entropy model. *Architectures and Mechanisms of Language Processing.* Edinburgh.

Schuchardt, K., Bockmann, A., Bornemann, G., & Maehler, C. (2013). Working memory functioning in children with learning disorders and specific language impairment. *Topics in language disorders, 33*(4), pp. 298-312.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal, 27*, pp. 379-423.

Van den Bos, K. P., Lutje Spelberg, H. C., Scheepstra, A. J., & De Vries, J. R. (1994). De klepel. Vorm A en N. Een test voor de leesvaardigheid van pseudowoorden. Nijmegen: Berkhout.

Van Ewijk, L. (2013). Word retrieval in acquired and developmental language disorders: a bit more on processing. Utrecht: LOT.

Wechsler, D. (1955). Wechsler adult intelligence scale. In *Nederlandstalige bewerking WAIS, 1970: Lisse: Swets & Zeitlinger.* New York: Psychological corporation.

9. Appendices

9.1 Entropy calculations

Low entropy 28*3 = 84 strings 7 | 7 syllables	
Syllable X	**Syllable Y**
4* kee	**4* muu**
4* joe	**4* goo**
4* daa	**4* lie**
4* puu	**4* vee**
4* teu	**4* reu**
4* hie	**4* saa**
4* foo	**4* sjoe**
H[bX] = H[7] = - ∑ x*logx = 2.8 H[XX] = H[7] = 2.8 H[XY] = H[7] = 2.8 H[Ye] = H[7] = 2.8 H[bXX] = H[7] = 2.8 H[XXY] = H[XYe] = H[7] = 2.8 H[bigram] = 2.8 H[trigram] = 2.8 H[total] = 2.8	

(Rădulescu, Wijnen, & Avrutin, 2015)

9.2 Training stimuli

Exposure 1/2/3

keekeemuu
joejoegoo
daadaalie
puupuuvee
teuteureu
hiehiesaa
foofoosjoe
keekeemuu
joejoegoo
daadaalie
puupuuvee
teuteureu
hiehiesaa
foofoosjoe
keekeemuu
joejoegoo
daadaalie
puupuuvee
teuteureu
hiehiesaa
foofoosjoe
keekeemuu
joejoegoo
daadaalie
puupuuvee
teuteureu
hiehiesaa
foofoosjoe

(Rădulescu, Wijnen, & Avrutin, 2014)

9.3 Test stimuli

Test 1		Test 2		Test 3	
type_1	daadaalie	type_1	hie_hie_saa	type_1	kee_kee_muu
type_2	poogaaroe	type_2	roe_nuu_nie	type_2	gaa_mie_suu
type_3	duu_duu_taa	type_3	zoe_zoe_voo	type_3	soo_soo_ruu
type_4	joe_daa_saa	type_4	puu_teu_muu	type_4	kee_foo_vee

(Rădulescu, Wijnen, & Avrutin, 2014)

Final test	
type_1	teuteureu
type_2	suunienuu
type_3	jiejiefeu
type_4	hiedaareu
type_1	joejoegoo
type_2	mienienuu
type_3	woewoesee
type_4	teupuugoo

(Rădulescu, Wijnen, & Avrutin, 2014)