Aalborg University Copenhagen

Semester:

Title: The Wor(l)d Builders *Designing for the Improvement of Middle School Pupils Writing Practices Utilizing Natural Language Processing and Didactics Strategies*

Aalborg University Copenhagen

Frederikskaj 12,

DK-2450 Copenhagen SV

Semester Coordinator:

Secretary:

Project Period: **01/02/2018 – 31/05/2018**

Semester Theme: **Master Thesis**

Supervisor(s): Lars Reng

Project group no.: NAN

Members:
Lukas Rasmus Nielsen
Samuel Kormind

Abstract:

This thesis revolves around the design of The Wor(l)d Builders. A design intended for motivating middle school pupils to write, and through this improve their writing capabilities. To that end, we research the fields of Didactics, Natural Language Processing and Game-Based Learning to find the most suitable methods that are to be included in the Design. The core of the design is to motivate through context. A context that is set through a narrative in a game world were the main interaction is writing sentences. These sentences are processed in Natural Language Processing to derive dependencies of words, and from these generate objects in a game world. The pupils will use didactic modelling strategies to learn how to write these sentences. Specifically, this thesis seeks to tests the proposed design to determine whether the methods included are viable in the context of The Wor(l)d Builders.
The design was implemented in Unity, and was tested on middle school and high school pupils ($n = 51$). Accuracy (67.4%) and Qualitative measures were used to test the design. In conclusion: the accuracy measure could easily be improved. Dependency parsing and didactic modelling are viable methods, while Word2Vec seems to fall short for the purpose of this thesis.

The Wor(l)d Builders

Designing for the Improvement of Middle School Pupils Writing Practices Utilizing Natural Language Processing & Didactic Strategies

Lukas R. Nielsen & Samuel Kormind

May 31, 2018

Acknowledgement

First and foremost, we would like to thank Teaching Associate Professor, Lars Reng, *Institute for Architecture and Media Technology, Aalborg University - Copenhagen* for his support and supervision, providing us with the feedback and encouragement to complete this project. Thank you for the supervision you provided. We hope to stay in contact in the future.

A special thanks to the head of Center for Social Practices and Cognition, Southern University of Denmark, Professor, ph.d. Gitte Rasmussen, for her great help with coordination, planning, and new ideas for improvement.

Furthermore, we would like to thank Research Associate Professor Eckhard Bick, *Institute of Language and Communication (SDU)*, and the rest of the team, who created the Visual Interactive Syntax Learning tool (VISL), for access to the VISL Dependency Parser.

Lastly, we would also like to thank our family and friends for support and motivation!

Reading Guide

We suggest that this thesis is read on a laptop or desktop computer. This is because we have used hyper references, enabling easy traversal from reference to section or figure and citation to Bibliography in the thesis. There are however one exception, which is references to the appendix which is included in a zip file on Digital Exam. The results of our tests are not transcribed for easy reading, but feel free to look through the data logs anyway.

Note that we use the terms *Educational Games* and *Game-Based Learning* interchangeably throughout this thesis, as we deem the two terms of similar meaning.

Contents

Chapter 1

Motivation

Throughout our years as Medialogy students, on both the Bachelor and Master semesters, we have worked with *meaning* and *understanding*, in various applications. All of which were to some degree meant as educational tools. Therefore, this thesis is based on our personal interest in educational technologies, and especially the application of such technologies in the field of game-based learning. After lengthy discussions with High schools-, middle schools- and University Teachers/Lectors, and aspiring teachers, most of whom are related to us as friends or family, we became aware of some issue regarding writing practices of pupils. It appeared that there was a lack of motivation to invest time in writing assignments, in the primary schools. After preliminary investigations concerning this issue, we found that an interesting solution could be to combine the field of game-based learning and natural language processing, as a method for encouraging pupils to improve their writing capabilities.

Chapter 2

Introduction

Throughout history, reading and writing has been one of the most important skills a person can accomplish [12]. In the modern world, the written language has a part in every aspect of life, and has a profound effect on one's ability to be a member of society.

Brok et al. (2015) argues that writing is taught in school, not only so pupils can learn to read and write the various academic and school related material, but to enhance pupils literacies. Writing is important, not only for the skill itself, but as a medium through which a person can note their thoughts, reflect on various topics, and communicate a message to a receiver. All of which are aspects, that constitutes a literate person. Writing as a literacy is a holistic integration, which means that writing is a part of reading, understanding and interpretation of the world [12]. Writing is taught from the earliest stages of school, and while most prominent in language classes, it is used in every subject as a means of obtaining and communicating knowledge. Unfortunately, issues arise once pupils have learned the initial spelling and grammatical competences required to formulate sentences: young pupils have a tendency to not only lose interest or to be discouraged in writing longer and coherent texts and sentences [18] [12], but also to be less engaged in classroom activities [8]. This phenomena can be caused by multiple factors, and no single factor has been confirmed. Instead the issue is most likely due to a combination of them. Outside of school, children use writing to chat with friends, communicate with family, post on social media, etc. In school, however, the purpose of writing is to learn, and this in itself is not always a motivational factor as most pupils feel that they write to receive grades or to please the teacher [12]. From a teaching perspective, the difficulty of encouraging and supporting pupils in written assignments is a reflection of the teachers ability to apply didactic strategies in their teaching method. With the introduction of modern media, as stated above, children use writing mostly as a medium for short bursts of communication, where shorter sentences and fewer words are a positive and practical necessity in

their attempt to reach, update, and expand their social circle. This contradicts the nature of the written assignment, which inherently is a tool for communicating messages, containing knowledge. Knowledge that can only be properly conveyed with some level of reflection, and explicit writing in order to understand the knowledge that is to be conveyed and not leave said message open for interpretation. Reflection is considered a metacognitive skill, that is often not fully developed in the minds of younger pupils, and could be the cause of their lack of interest in writing [18] [41]. Young pupils fail to understand, or have yet to experience, what a good and coherent sentence and text can do when writing, either for themselves or for others. Not considering the pupils lack of interest, the causality of this issue could very well be the applied teaching strategies [18] [27]. Such strategies include methods of stating goals for a given assignment, helping the pupils through the use of modeling techniques, and providing proper feedback and evaluation (see the Writing & Didactics section). Feedback in school is mostly given through summative evaluation, even though formative evaluation is commonly understood as yielding better results [12]. Formative evaluation is a continuous method of evaluation which simply takes up too much of the teachers time in the classroom. The field of computer science could help alleviate the teachers workload, regarding feedback and evaluation. One of the core strengths of a computer application is its ability to provide immediate, individual, feedback to pupils. Furthermore, it has been found that to indulge younger pupils in learning, video games can be an engaging tools for conveying the learning content, and seemingly hereby a positive effect on the learning outcome [23], [44]. The idea of game-based learning sprung from the notion that the process of learning should be engaging as well as educational. Through video games the pupil becomes an actor that, through the virtual world, can make meaningful actions that affect the course of the unfolding narrative. Such agency has been shown to have a positive impact on pupils ability to retain information as well as learning [4].

Game-based learning are developed for the specific purpose of education. However, it is the core concept of video games that, if applied correctly, can bring a positive outcome to education [36],[23],[20],[33],[44],[3],[14]. The notion of video games' ability to provide immediate, individual feedback serves as the foundation for this project. For a given computer application to be able to provide immediate feedback, without any external evaluations (e.g. Teachers) it must be able to decipher what the pupils write. In other words, the computer application must be taught how to read, in order to evaluate how well a pupil writes.

The field of *Natural Language Processing* (NLP) is specifically that; teaching computers the written language [13]. Through the employment of computational techniques researchers use NLP to understand, learn and produce human language. NLP has become almost ubiquitous in our society. It is what enables Apple's Siri to communicate with the consumer, how

Google Translate translates between languages and how advertising on social media platforms such as Facebook work (i.e. through text mining). For the specific purpose of analyzing written texts, the utilization of machine learning (ML) techniques, has made NLP quite sufficient in interpreting bodies of texts. For an application to be able to understand a written text, a variety of NLP methods can be put to use; tokenization (i.e. splitting a corpus into sentences or words), part-of-speech tags (i.e. determining whether a word is a noun, verb, adjective etc.), stemming words (e.g. deriving the word 'token' from 'token-ization') [10]. These methods need to be combined in such fashion that a system can be trained, through ML, to understand that in a sentence such as "The man enjoyed watching the beautiful waves of the ocean." it is indeed the waves that are deemed beautiful and not the ocean, nor the man.

Specifically, we seek to design a writing application called *The Wor(l)d Builders* that, through the utilization of didactic strategies accompanied with appropriate techniques from the fields of educational games and natural language processing, can be used to provide pupils with a tool through which they can, and will want to, improve their writing capabilities. We believe that such an application is best suited for middle school children in the 5th to 6th grade based on the learning goals set by the Danish ministry of education [48]. Throughout the 4th to 6th grade, the pupils in Danish middle school have to learn the fundamentals of writing, and are by that time in their education, being taught the importance of proper content and fluency. These skills are expected to be fully acquired by the pupils once they reach 7-9th grade where they are required to formulate longer and coherent texts.
This thesis will revolve around the design of *The Wor(l)d Builders* application. The focus will lie on designing and implementing didactic strategies, and Natural Language Processing tools. Specifically, we raise the following problem statement:
How can we design The Wor(l)d Builders *to interpret the written input of Danish middle school pupils' and provide motivational feedback?*

To answer this problem statement, we will analyze research within the fields of Natural Language Processing, Educational Games, Didactic strategies and writing, eventually defining which NLP methods, and didactic strategies will be included in the design of *The Wor(l)d Builders*.

Chapter 3

Research Analysis

Our research will be acquired mainly through the Aalborg University library, and the umbrella database EbscoHost, covering Academic Search, Premier, J-Stor, Proquest Research Library, Scopus and Web of Science. To find relevant research we used key words/phrases such as: *Natural Language Processing, NLP, Word2Vec, Part of Speech Tagging, POS, Intelligent Tutoring System, ITS, Educational Games, Writing, Composition, Serious Games, Game-based Learning.* The field of Natural Language Processing is completely new to the authors, therefore courses on the subject will be taken via internet resources, such as Standford's lectures on Natural Language Processing.

In the following sections relevant research which has been found using the key words and phrases mentioned above, will be presented and analyzed eventually culminating in a set of design requirements.

3.1 Writing & Didactics

Writing in school is usually done through the pupil taking notes, answering questions and writing lesser texts. The latter is often the objective of "larger" assignments, where the teacher will provide a goal for a written hand-in, and thus pupils will write and receive feedback and/or grades based on their performance. In most scenarios, the evaluation of a written assignment will be given once the product is finished, and will be based on the content and grammatical capabilities displayed in the assignment. This is called *product writing*, and refers to the evaluation of an assignment as a finished product [12]. In product writing, teachers rarely provide pupils with more than a short '*well done*' note and a grade. While this definitely has its place in education, studies suggest that process writing is more motivating, and tends to increase the learning outcome of an assignment [12] [27]. In process writing, a writing assignment will be split into three stages: *pre-writing, writing* and *post-writing.* In process writing, there is an active

helper and participant in all three stages, providing guidance and feedback throughout the process. This role is usually fulfilled by the teacher. The pre-writing stage concerns reflection, planning, and preparing for the given assignment. This leads to the writing stage, where pupils receive continuous feedback from their teacher during the process. Finally the post-writing stage, where the product and process is evaluated. Depending on the given assignment, these three stages might be reiterated multiple times, enhancing the pupils experience for each iteration.

3.1.1 Pre-writing

Pre-writing is a stage that will always take place during a school assignment, but one of the purposes of process writing is to fully develop this stage, with the teacher as a guide. Properly motivating pupils to write a given assignment, is just as important as giving the assignment itself. More energy should be used on stating clear goals for writing assignments, so that pupils have a purpose with their writing [12]. When a teacher provides the class with a written assignment, Brok et al. (2015) suggests the following principles, to make sure that the given assignment is understood correctly, and encourages writing:

Writing purpose Why should the text be written?

Educational purpose What will the pupil gain from writing the text?

Receiver Who will read and understand the text?

Content What will the text concern?

Action What will the text be used for?

[Brok. et al, 2015, translated from danish]

Writing purpose, educational purpose, receiver, content and *action* are keywords that can be used to structure learning material, given teachers a framework to utilize these concepts and provide context to the pupils. The above mentioned principles can be seen as metacognitive (i.e. thinking about thinking - planning) tools to enhance pupils reflection of their work. Reflection is an essential part of the pre-writing stage, and serves as a motivator for a given assignment [18] [34] [41]. Teaching pupils to reflect upon *how* and *why* they should write an assignment, can help motivating them in completing the task. Smidt et al. (2013) proposes different reflective techniques, that can be related to the principles presented by Brok et al. (2015). *The ten thesis*, as Smidt et al. (2013) calls them, each serves as a tool for reflection, planning, assessment and evaluation, that teachers of every subject

can utilize during process writing. [43] argues for the importance of discussing the purpose of an assignment with the classroom (*writing purpose and educational purpose*), create an understanding of the written language for each type of assignment (receiver, content and action), and provide clear goals related to the evaluation criteria (writing purpose, educational purpose, receiver, content and action). The reflective techniques and principles can help pupils understand the purpose behind their writing assignments, and motivate them to do more than simply '*write to please the teacher or to receive grades*'.
It is important to note that instructions for a given assignment are easier to comprehend when given on text, than when presented orally. Furthermore, it is important that the assignment is always present and accessible for the pupil, so they can look it up whenever they are in doubt.

3.1.2 Writing

The essential aspect of the writing stage in process writing, is the teacher's role as a constant guide, providing tools and feedback throughout the pupils writing process. A common issue, other than feedback, that occurs among pupils during a writing process, is the lack of knowledge and inspiration as to how such an assignment is written.

3.1.2.1 Modeling

Modeling techniques are practical methods of guiding one's pupils. With modeling techniques, a teacher can provide cases in which a similar assignment is applied, and text based examples where pupils can experience the language and form of such work [12]. A teacher can begin an assignment by writing the first couple of lines for each section of the text in front of the class, while explaining the purpose of the sentences. Furthermore, they can write a similar text on the whiteboard, answering questions related to the assignment (applicable during lesser assignments). Modeling techniques should be seen as references and support, which a pupil can use throughout the writing process, motivating them to complete their work. Agus and Winiharti (2011) conducted a study on the effect of storybook reading to aid pupils in writing assignments. Their method is a great example of a modeling technique, where pupils learn through examples. Every day for weeks, the teacher would read a chapter from a storybook aloud, and proceed to an open classroom discussion concerning the design of characters, story, setting, etc. Every day the pupils would then continue working on their own story based on what they learned. The teacher then applied another modeling technique, where students were taught to visualize their stories through drawings. They were told to draw their story's beginning, middle, and end in different colours, and then put them together. Through

this method, pupils could clearly see how a story is structured and how it fits together.

3.1.2.2 Writing Motivation

It is important to note, that every pupil has different methods of writing and visualizing their strategies, and every pupil has an individual feeling and opinion towards writing. Shapira (2007) conducted a study on the different writing profiles of six grade pupils. The purpose of their study was to see how each individual approaches writing differently, and to create more awareness of this among teachers [41]. Their writing profiles were split into four clusters: *Metacognitive, Cognitive, Social,* and *Affective*. They conclude that six graders have yet to develop their metacognitive and cognitive strategies, and thus social and affective profiles were much more common. Social refers to the relationship between a pupil and their teacher, and his/her peers, and concerns feedback and help. Affective refers to one's feelings towards writing, whether it is positive or negative. The study found that pupils with a positive affect towards writing, where more prone to create a better writing output, as they were generally more fond of the task. Pupils with a negative affect towards writing would greatly benefit from better use of the social profile, as help from their fellow peers and teacher can increase their motivation.

3.1.3 Post-writing

The post-writing stage in process writing is generally split into two parts. The first takes place after the initial writing stage, and concerns the teachers immediate feedback on the content of the assignment. This stage might lead to another iteration of the writing stage, or it can lead to the second post-writing stage, where the writing process is evaluated as a whole, and pupils are given a final verdict of their work. There are two methods of evaluation which a teacher can utilize: Summative evaluation and Formative evaluation [12]. Summative refers to evaluation methods, such as grades, comments, and check marks. These are given at the end of an assignment or exam, and are mostly used as a method for teachers to keep track of the educational progress of their pupils. Summative evaluation is, unfortunately, the most common form of evaluation. This is not to say that summative evaluation is solely negative, as it still serves as a practical tool once an assignment is finished. Many teachers use summative evaluation, as it is less time consuming, and are based on the tradition of teacher control [12]. Formative evaluation, or continuous evaluation, refers to when feedback is given throughout every stage in an assignment. Teachers report that applying formative evaluation in practice is not always feasible due to the quite extensive amount of time

it requires (each individual must have unique feedback). This is of course a problem, due to the benefits, for the pupils, of receiving such feedback. Providing pupils with formative evaluation also revolves around the creation of self evaluation techniques in order for pupils to notice where they went wrong, and what they can do to improve. Svanes and Skagen (2016) argues that the formative evaluation should be tailored to each individual pupil, as pupils should receive feedback based on their unique character and knowledge of the subject. If the pupils receive tailored feedback, self-evaluation may also occur more often, as they get insight in the concrete mistakes they made, and how to improve upon them. Self-evaluation can be understood as reflective techniques (see Pre-writing section) and revolves around the goals and purpose of the assignment. Therefore, it is important that the teacher has the necessary tools to ensure that these techniques are present throughout process writing.

Smidt et al. (2013) proposes that teachers should provide pupils with assignments that have a purpose, after being finished. This could for instance be writing a Wikipedia page, that serves the purpose of communicating knowledge to other people. He refers to this as '*setting texts in motion*', and argues that too often, school assignments simply end up in the teacher's desk. Teachers can provide a function for a given assignment, such as using them for a school exhibition, reading them aloud to each other, or simply hanging them on the classroom walls. By utilizing this method the teachers can show that the assignments have real world impact, and in fact matter to an extent that exceeds grades and making the teacher happy. A notion that should encourage pupils in their writing.

3.1.4 Didactic Perspective

In conclusion we see a relevance for the following key points in didactic strategies:

- Writing should be done as a process, where the pupil is constantly met with help and feedback, tailored to his/her needs.
- The pupil should be taught to reflect on, and plan, their writing, and have the opportunity to do so.
- It is important that the goals and objectives of a written assignment is clearly stated. Furthermore, the goals and objectives should be easily accessible.
- The pupil must be taught the purpose of their writing assignment, in relation to *writing purpose, educational purpose, receiver, content* and *action*. This will greatly increase encouragement and motivation, as the written assignment becomes more than simply homework.

- Modeling techniques, which will provide inspiration through examples of similar texts. Consequently yielding a better understanding for the pupil.
- Positive and negative affect towards writing, have an impact on the willingness to participate in assignments. Integrating methods that cater to social interaction between pupils, e.g. peer feedback and help, could help alleviate this issue.
- Formative evaluation should be used throughout the writing process. This includes encouraging pupils to utilize self-evaluation techniques.
- The written assignment should have a purpose beyond completion, meaning that it serves some purpose of handing it in.

One could argue that it would require a superhuman teacher to successfully apply all the key points in their teachings, however it could be made possible with the aid of a computer application. An application could serve as tool, which provides feedback to pupils during their writing process, while fulfilling other functions, such as writing purpose and modeling techniques. To that end, Natural Language Processing is required (see Introduction section). The following chapter will expand the scope of educational games to E-learning, because there is, to the best of our knowledge, a scarcity of application utilizing Natural Language Processing, that fully adheres to the concepts of educational games. Furthermore, the Danish Department for education, calls for the utilization of IT systems to educate pupils [49].
Therefore, we will try to encapsulate the essential aspects of educational games, or game-based learning, and Natural Language Processing in the following chapter for the purpose of utilizing different prominent techniques in the *The Wor(l)d Builders* application.

3.2 E-Learning

In the field of electronic learning (E-Learning) there are two fields of research that are especially relevant for this thesis, namely Intelligent Tutoring Systems and Game-Based Learning.

3.2.1 Intelligent Tutoring Systems

Intelligent Tutoring Systems (ITS) are applications that utilize Natural Language Processing for tutoring purposes. A number of these have shown promising results regarding evaluation, assessment and feedback in connection with written assignments.
Roscoe et al. (2013) have created an ITS called The Writing Pal (W-Pal) for the purpose of improving students writing efficiency. Roscoe et al. (2013)

does this through evaluation of essays, written by the students, and mini games. The evaluations are based on a variety of natural language processing tools:

Lemmatizer Finding the lemma of a word (e.g. '*good*', from '*better*')

Syntactic Parser Recognizing an input string, and assigning a syntactic structure to it.

Lexical Database For example WordNet

Rhetorical Analyzer Recognizing the Rhetorical strategies

Latent Semantic Analysis Relating content of e.g. two documents to one another.

The evaluation is not merely based on scoring results, but is also used to detect whether or not students adhere to different writing strategies. In the mini games no writing occurs. Here the students must select a correct answer among several options of examples of e.g. essay introductions and attention-grabbing techniques. However, the content of the mini games and the different options that the pupil must select from, have no relation to one another. The mechanisms of the games seem forced upon the subject that the pupil is to learn about. This lack of contextualization may result in the game aspect not contributing to an understanding of the subject, other than the assignment might be more enjoyable for the pupil. Furthermore, the writing aspect of W-Pal is not focused on contextualizing the assignments, but rather to give them, the pupils, a holistic rating of their essay (e.g. poor or great) and feedback on how to improve their writing, in general terms of proper essay structure. Wilson et al. (2017) attempted to use natural language processing with the purpose of creating a tool for automated formative writing assessment. Their system utilizes the levels of language framework as the base of the assessment. The system, based on the framework, allows them to not only assess the content of the written assignment, but analyze deeper levels such as grammar and spelling. This was done through the use of the *Coh-Metrix*, which is a tool developed to assess different aspects of written input. Similar to the W-Pal system, the Coh-Metrix uses *lexicons, pattern classifiers, part-of-speech taggers, syntactic parsers, shallow semantic interpreters, latent semantic analysis,* and *other computational linguistics methods and techniques* [50]. Through two consecutive experiments, test participants of grades 6-8 (American middle-school), would access an online writing prompt, where they would be tasked with different writing assignments. Participants had an hour to finish the assignment, and would then be assessed based on their words, sentences and discourse in accordance with the levels of language framework. This

assessment provide the pupils with knowledge which they can use for self-evaluation, as they are told in which of the levels (i.e. level of language) they are lacking, or have learned the skills that are required of them. Furthermore, the teacher can use the assessments to provide further feedback and guidance to each individual student in correlation with their needs. Wilson et al. (2017) stresses that automated writing assessment tools have already been proven to be efficient and reliable, but through the use of the framework, they successfully developed algorithms that can notice and assess deeper individual differences in student writing.

These are two examples of successful intelligent tutoring systems, which proves that a computer application can indeed be used for real time assessment. While this may be of practical use for teachers, it seemingly changes little when it comes to encouraging pupils. One could argue that from the pupils point of view, they are still simply writing an assignment, even if the formative evaluation can improve their learning outcome. With the exception of the W-Pal mini games, these two examples are not meant to be entertaining. If we seek to encourage pupils to improve their writing capabilities, then we must do more than just assess their writing. The intelligent tutoring systems can serve as inspiration in their use of natural language processing, and as proof that the current state of NLP is in fact quite able to evaluate written texts, but when attempting to encourage pupils, they must feel as if their writing has a purpose beyond simply learning. The field of educational games and game based learning can aid us in providing entertainment, goals and purpose to the writing application.

3.2.2 Game-Based Learning

Over the past couple of decades, many researchers have studied the utilization of games as a means for education. Games can be a resourceful tool because young people, in particular, spend many hours engaging themselves in games and learning the rules of the systems. A system that may be structured around learning material. [11] and [14] argues that: "...games are productive in helping students apply, synthesize, and think critically about what they learn through active and social participation.". In the early years of childhood, children use play to learn the necessary skills that is required to be a well-developed individual in society [19], regarding social, emotional as well as academic competences [17]. Video games provide a kind of play from which one does not necessarily learn social competences that are directly transferable to the physical realm of man. However, in online play, such as in the Massive Multiplayer Online Role Playing Game (MMORPG) World of Warcraft, players need to learn, not only the systems of the game, but also how to interact with other players [14]. This interaction ranges from everything between aiding a fellow player in need, to how to properly utilize the messaging system of the game (incl. learning the game vocabulary). In

fact, playing games such as World Of Warcraft have shown to improve the English skills of children [25].

Writing in school is generally taught in such fashion that the texts produced by the pupils are decontextualized from the topics they are writing about [14]. Contextualization is important for some learners because we, as humans, learn quite differently. Some learn through memory, while others need to get their 'hands dirty' (e.g. experiencing the phenomenon they are writing about) [17]. Through computer games it is quite possible to create assignments that are based on occurrences within the simulated worlds, meaning that the assignments can be produced from or, preferably, within context. Barab et al. (2012) presents a theoretical framework called Transformational Play. This framework is centered around designing game-based curriculum. Its intent is to make students protagonists who have to make sense of the world they are situated in (i.e. a game world). They make sense of this world by utilizing their knowledge, skills and concept of educational content. This utilization must have an impact on the world they are inhabiting, to foster a better understanding of the educational content [5]. There are three important aspects of Transformational Play;

Person which is linked to intentionality

Content which is linked to legitimacy

Context which is linked to consequentiality

Barab et al. (2012) had a teacher divide three 7th grade classrooms (n=63) into a game-based group and a story-based group, to ensure that the skill level of both groups did not differ. The students, being positioned as actors in the narrative, were asked to write a persuasive text. Their task was to either write persuasively from the perspective of a character or to convince a character in the story. The participants of the game-based group were actors within an video-game application called *Plague: Modern Prometheus*, a fiction based on Mary Shelley's *Frankenstein*. They were provided with a tool to capture quotes from conversations with characters within the game, in order to reinforce the arguments in their writings. Depending on the students decisions and interactions within the game, the narrative would unfold differently, hereby making sure that the actions, that the students made had consequences. The story-based group, who did not play a game, worked with the narrative in the novel *The Clay Marble*. The students ability to write persuasively was reinforced by the teacher providing lectures and presentations on said topic. Due to the nature of a classic novel, the participants of the story-based group did not experience consequentiality. They were, however, asked to write from the perspective of different characters in the novel, to affect the outcome of the story. This was of course only

in an imaginary sense, since there were no visualization, or otherwise illustrated means, of their writings. Throughout the study multiple tests were conducted over the course of 12 classroom periods; Observations, Pretest-posttest measures, Engagement measures and Researcher interviews. The results of the pretest-posttest showed that the effect size of learning gains of the two groups were of significance (story-based = 1.22, game-based = 1.83), but that the game-based groups had significantly higher effect size on learning gains. Due to the nature of their tests, some bias may occur as the two groups are presented with different stories, because one might be more intriguing and thus, more engaging than the other. It is, however, quite difficult to tell the same story through games and books. One of the mediums is bound to fall short due to agency and the expectations of the medium (e.g. in games, participants might assume they will have more control than in a book). In terms of engagement [5] measures flow [15], which has been linked to engagement, through a 10-point likert-type measure. The results showed that the game-based group experience higher levels of flow. Overall, the pupils in the game-based group showed more enjoyment, commitment to the task and understanding of the content. These results are yet another example to amplify the potential of game-based approaches as means of education. Below is a list of key elements, derived from this section.

Actors The pupils must become actors, who have an impact on the game world to contextualize the learning material

Social Interaction seem to encourage learning.

Information Storage The pupils ability to capture information in [5] seems like a good method to pertain information.

Purpose Having a clear goal, seems like a motivational factor (be it, convincing a fictional character, or another player to help you)

Having pupils play a game does not necessarily guarantee a high learning outcome. The game must be created with education in mind, or as Colby and Colby (2008), having the curriculum center around the content of the game. An important aspect of the utilization of game-based learning is instructional techniques. How do we, as developers, properly present the content, feedback, evaluation etc. to the pupil, for them to have the best possible learning outcome?

3.2.3 Instructional Techniques

Ferdig and Winn (2009) presents the Design, Play and Experience (DPE) framework as can be seen in figure 3.1. The figure portrays the relationship between the designers intentions and the players experience, and includes the

considerations, that a developer must make through three stages of creating an educational game.

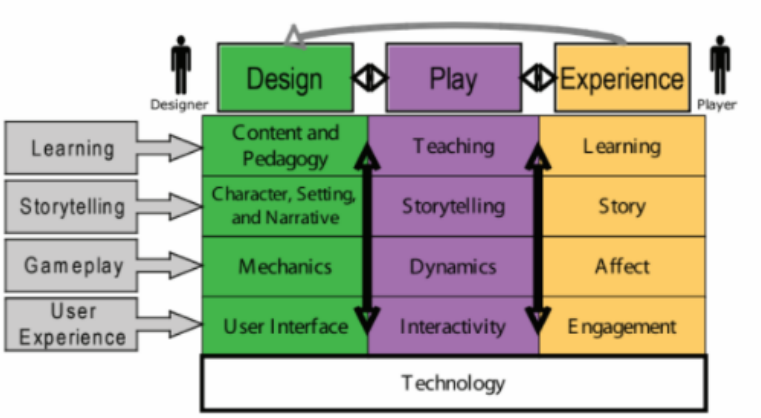

Figure 3.1: The DPE framework [22]

The design stage refers to the purpose and development of each layer. The play stage is the '*physical*' game itself: what the design will lead to in terms of gameplay and mechanics. The experience stage refers to the concept of player gain: which affect the different layers will have on the player. The layers are the four different aspects of an educational game (according to Ferdig and Winn (2009)). They are each represented through the three stages, and serve as guidelines for development. Finally, Ferdig and Winn (2009) presents technology as a fifth layer, and clarifies that all design decisions should be made with a designated technical platform in mind. When one attempts to design for the learning layer, with the three stages in mind, Ferdig and Winn (2009) suggests to look for instructional techniques to aid in the didactic design. Wouters and van Oostendorp (2017) conducted meta-analysis of instructional techniques for the development of serious- and educational games:

Assessment and adaptivity Adapting the game to the learners abilities through real-time assessment

Level of realism The auditory and visual representation in the game.

Narration-based techniques All the narrative techniques used to further story and learning

Feedback The given feedback to the learners actions. Can be displayed as corrective (right or wrong) or explanatory (why correct or not)

Self-explanation and reflection Learners are encouraged (or "forced") to reflect upon their answers.

Collaboration and competition The social aspect of the game. Peer discussions, playing together or competing.

Content Integration Learning content is embedded into game mechanics, such that learning happens simply by playing.

Context Integration Using the game as part of a larger lesson, where the game is only a part of the learning material.

Modeling Help or indications as how to solve problems in the game. Can be verbal, animated or graphical.

The instructional techniques, if implemented correctly, should enhance the effectiveness and motivational aspect of an educational game [51]. The *modeling and feedback* techniques, much related to the strategies stated in the Writing & Didactics section, proved to be effective in supporting learners [51] . These can be seen as part of the Play stage , and should be implemented as an educational tool in the learning layer (fig. 3.1). *Assessment and adaptivity* relates to the learning layer as well, and has the potential of becoming a powerful learning tool, according to [51]. Proper usage of this technique can turn out be difficult to implement, but does not make adaptivity any less essential, as a learning environment must closely match a pupils skill level, such that the tasks are neither too difficult nor too easy [40]. Balancing these factors is essential if one seeks to promote a state of flow in the player [15]. This means that some form of adaptivity should be present, through either (real time) assessment or a manual scaling. Roscoe et al. (2013) argue that the perceived helpfulness of an educational game can positively affect enjoyment, where helpfulness refers to the help mechanics present in the game [39]. Perceived helpfulness can be created through subtle hints and aid throughout the game. *Content integration* is another powerful tool that mixes the learning, storytelling and gameplay layer (fig. 3.1). Through embedding the learning material in the game mechanics and narrative, one can ensure that the player will have to learn in order to progress. With said narrative being interesting and the game mechanics entertaining, the player might even enjoy learning. The narration-based techniques proved to directly affect motivation of the players but not affect learning. Didactic research seems to indicate a direct correlation between motivation and learning [12], thus one could argue that a direct effect on motivation would indirectly affect learning. *Context integration* creates flexibility around an educational game, meaning that it has the potential to be used throughout different lessons. A game with the purpose of encouraging pupils to write, could be used as part of any larger writing assignments. This instructional technique can also relate to the concept of setting texts in motion (see Writing & Didactics), where pupils can use the game as a part of a larger context. *Collaboration and competition* are present in the gameplay layer (fig. 3.1),

but can be used to increase the motivation of pupils, especially those with a negative affect towards writing (see Writing & Didactics section). *Level of realism* yield more positive learning experiences, when simple and cartoon like graphics are utilized. Realistic graphics can be distracting, and remove focus from the actual learning experience [51]. Furthermore, creating realistic graphics is time consuming, and seems beyond the scope of this project. *Self-explanation and reflection* is not directly present in the DPE framework, but one could argue that the sum of all the layers could promote such metacognitive skills (fig. 3.1). As mentioned several times in the Writing & Didactics section, reflection is essential for an improved learning experience, but moments of reflection are often rare in educational games [40]. S. Paras and Bizzocchi (2005) argue that the best gameplay experience is had, once the player achieves the state of flow, but that no reflection happens during the flow state. They emphasizes that educational games should allow for moments of reflection, or as Wouters and van Oostendorp (2017) call it: '*force*' players to reflect upon their learning. *The user experience* layer of the DPE framework, encompasses everything that the user will interact with during a playthrough. Ferdig and Winn (2009) states that the user experience layer leads to engagement in the player, but we argue that the entire design stage of the DPE framework should, if implemented correctly, provide engagement. We base this on the assumption that the design stage of an educational game can be understood as developing a learning environment. A learning environment refers to a structured setting with the purpose of learning [6]. A proper learning environment should promote engagement in pupils. Norman (1993) states that a successful learning environment should: '*Provide a sense of direct engagement, producing the feeling of directly experiencing the environment, directly working on the task*'. Furthermore, a learning environment should promote a proper user experience, or in other words: '*Provide appropriate tools that fit the user and task so well that they aid and do not distract. Avoid distractions and disruptions that intervene and destroy the subjective experience*' [31]. The Graphical User Interface (GUI) should be subtle and unintrusive. It should be minimalistic and only present when necessary. Ferdig and Winn (2009) emphasizes that the experience stage can be very different depending on the player's cognitive, social, and cultural background. A developer should define and develop their game with their target audience in mind. In middle school, both boys and girls seem to prefer games with explorative content, and both enjoy games that involve problem solving [24]. The majority of middle school boys prefer active and strategic games, that are centered around goal-oriented objectives. Girls prefer games that enable creativity. Furthermore, girls prefer games that allow for social interaction, are non-violent and have an interesting narrative [32]. When it comes to characters and setting, both middle school boys and girls prefer characters to be similar to themselves (i.e. skin colour and characters that are a slightly older than themselves). Both boys and

girls prefer a male character to be of muscular build. Boys prefer a female character to be fit, while girls prefer female characters as slight. When asked about the purpose of their actions in a game, the majority of boys and girls stated that they wanted to "save all living things on the planet". When stuck in a game, the majority preferred the self-help method. This refers to "...either methodically attempting to solve the problem or discovering the answer through trial and error" [24]. These findings clearly portray the similarities and differences between middle school children's game preferences. We can use these findings when designing and developing the educational game for the target audience.

3.2.3.1 Instructional Technique Key Points

In this chapter we analysed the DPE framework by Ferdig and Winn (2009) and the instructional techniques that [51] found most promising through meta analysis. Below we derive a list of key points from the Instructional Techniques section, which could potentially be used in the design of *The Wor(l)d Builders*:

Modeling and feedback The findings in [51] are in accordance with the didactic strategies found in the Writing & Didactics section, and seem essential to the design of *The Wor(l)d Builders*.

Adaptivity While the application should be designed for different skill levels, we do not deem it necessary in the first iterations of the design, since we need to establish what is, and what is not difficult.

Learning Material The learning material should be embedded in the design of the narrative and the game mechanics.

Narration The narrative should be interesting to increase pupil motivation.

Context Integration goes beyond the application. We might write a narrative that can be suited for multiple classes, e.g. Biology and Danish, hereby giving the application multiple meanings.

Social interaction The game should encourage social interaction, both between classmates, and between pupil and teacher, to help the less fortunate.

Graphics should be simple and cartoon like.

Reflection The game should require the player to reflect upon their learning.

The Graphical User Interface Should create a pleasant user experience, and be devoid of distractions.

Gameplay The game should be centered around exploration and problem solving. It should have elements of active play and creative play.

Avatar Instead of presenting characters in game based on middle school childrens' preferences, we could simply use first person characters only. This would be easier to develop, and we would avoid any dissatisfaction's among pupils.

Objective The objective of the game should concern *'saving all life on the planet'*

3.3 Natural Language Processing

In Natural Language Processing (NLP) there is a multitude of sub fields. Whether it is the generalization of the speech patterns of conversational agents [7], aiding pupils in evidence based writing [35] or identifying linguistic features of stigmatized groups [37] the underlying "*... goal of Linguistic Science is to characterize and explain linguistics observations...*" [28] that occur all around us, in our everyday life.

To serve the purpose of this thesis, *The Wor(l)d Builders*' NLP system must be able to characterize word classes and sentence composition, in order to use and understand pupils' input. Take the following sentence: *"In the center of the valley an old, yet beautiful, tree grows on the bank of a small river."* This sentence has four objects of focus, namely valley, tree, bank and river. These are nouns, and represent the elements that make up the sentence. Generally, the system must recognize nouns as objects, but there are some exceptions such as the word center in the sentence above. Center is also a noun, but is instead used as a descriptive feature of the sentence similar to adjectives and adverbs. In cases such as this, it is important that the system has been taught compositional knowledge for it to recognize the relationship between nouns, when such nouns are separated by words of classes such as preposition and determiner (e.g. "of the" in the above sentence).

3.3.1 Word Ambiguity

Words can have multiple meanings, depending on the context in which they appear (i.e. the semantic meaning of a word). Take for instance the word

'*bank*' from the sentence presented in the Natural Language Processing section. In English (and Danish) the word has multiple meanings, and senses:

- English Definitions from Cambridge
  - '*Bank*' an organization where people can invest and borrow money etc.
  - '*Bank*' In gambling, the bank is money that belongs to the owner and can be won by the players.
  - '*Bank*' Sloping raised land, especially along the side of a river.

In NLP, a task could be to select the correct meaning from a set of dictionary entries, depending on the context that the word is given in. To train the model to be able to perform such disambiguation tasks [28] discuss a variety of supervised machine learning (i.e. treating it as a classification task) and unsupervised learning (i.e. treating it as a clustering task) approaches to the problem. [28] argues that unsupervised learning is not possible if the purpose of the model is to tag a given word with semantic meaning. It is however, possible to make '*context clusters*' of an ambiguous word, without tagging the clusters. In the case of an application such as *The Wor(l)d Builders*, unsupervised learning does not seem feasible. The model should be able to determine the semantic meaning of an ambiguous word (and give an appropriate response), and not only determine whether or not the word in question belongs to the same cluster, as the same ambiguous word, in another context. With supervised learning there is of course a need for manually labelling the data that the program is to be trained on. Considering the diversity of a words usages (e.g. '*Bank*'), this could be a very long and tedious task. Therefore we will need to find lexical resources with labelled data, if we are to train a model based on a supervised approach.

3.3.2 Vector Representation of Words

Adjectives and adverbs are, as mentioned, the descriptive feature of each sentence. These will often determine the visual output that a pupil will experience. The system must be taught how to interpret these words for it to produce the proper output. Colours are typical adjectives and adverbs, which are also simple to read. In cases such as the *"In the center of..."* sentence above there are adjectives that provide a description of the objects/nouns, but adjectives can have synonyms that are similar, though not entirely the same.
Take the following example: "*"In the center of the valley, an ancient, yet magnificent, tree grows on the bank of a modest river."*. The adjectives have been replaced with synonyms and the general meaning of the sentence remains the same. Some would argue that the exchanged adjectives do not

adhere to the same exact meaning, such as the difference between old and ancient, where the latter often refers to something that is "very old" compared to "just" old. When training a system to recognize and understand adjectives, the system would most likely interpret these as being the same or similar. Such synonyms would be perfectly fine for the sake of a prototype, but one must be aware of situations where the meaning can differ extensively.

Traditionally, most NLP approaches used atomic representations of words. The word '*Berlin*' might have been represented as *Id01* and the word '*Copenhagen*' as *Id13*. When a model, using such symbols, was trained it would learn all the features of '*Berlin*', and then all the features of '*Copenhagen*' without realizing the relationship that exists between the two (i.e. they are both cities, they belong to a Country and are both their respective countries Capital).[29] presented a method that represents words as vectors, utilizing a skip-gram model, called *Word2Vec*. The idea of representing words as vectors was not entirely new at the time, but the dimensionality reduction that followed was quite useful when compared to other vector representations of words, such as one-hot encoding. Take the following sentence: '*one hot encoding words*'. To represent each word in this sentence as a one-hot encoded vector, they would be allocated as the integer 1 in a vector of zeros, that has the same length as the corpus of text:

$$One\ [1\ 0\ 0\ 0]$$

$$hot\ [0\ 1\ 0\ 0]$$

$$encoding\ [0\ 0\ 1\ 0]$$

$$words\ [0\ 0\ 0\ 1]$$

such representations can become quite extensive as Danish for example has 200.000 - 300.000 words, and grows by the thousands every year. Furthermore there is no obvious semantic relationship between each vector, unless the data is explicitly labelled with categories describing such relationship.
The Word2Vec model on the other hand, can represent vectors with e.g. 300 features, as opposed to the size of the entire vocabulary and is able to pertain semantic, as well as syntactic, features of words in a continuous vector space. This means that words that have similar features, such as '*Berlin*' and '*Copenhagen*' will be embedded near each other in said vector space. Mikolov et al. (2013) describes how one might perform algebraic operations on vectors in the model, operations that would reveal semantic similarities. One might train a Word2Vec model, and find that:

$$v_{[Berlin]} - v_{[Germany]} + v_{[Denmark]} = v_{[Copenhagen]}$$

The model has been improved upon, enabling it to make better vector representations of words, and decreasing the time it takes to train it [30].

Word2Vec seem like a viable means to give pupils more freedom to express themselves, which is why we seek to train and apply such a model to *The Wor(l)d Builders* application, to the end of deeming whether or not it is a feasible model for our design purpose.

3.3.3 Part of Speech

Part of speech (POS) tagging, is the process of allocating words, given a sentence, to their respective classes. In order to train such POS taggers, label data is needed, as an unsupervised approach is not possible. There are a variety of tools that are freely available, which have been trained on labelled data. Some of which are quite sufficient in POS tagging Danish words in sentences. One of these are described in [2], and is called Polyglot. The model was trained on more than a 100 languages. They found that their algorithm performed better than, or as good as many state of the art systems. Their model was trained in different stages. First, the model was trained for word embeddings (see Vector Representation of Words section). The data used for training was acquired from Wikipedia. When they found the model had peaked on performance, they trained a POS tagger on languages were they were able to find annotated data (one of them being Danish). Their findings are now freely available, and Polyglot is accessible through Python. Bird et al. (2009) Has created a tool called NLTK, which is also freely avaiable. [10] is an extensive read, about the different methods that the toolkit provides, and how to use them. We will through the first iterations of our design phase deem which of the two, if not both, is best suited for the purpose of this thesis.

3.3.4 Sentence Composition

It is important for our system to know the relationship between adjectives and nouns, such that it understands when adjectives refers to a specific noun, and not other nouns in the same sentence. The same applies to verbs. These are generally seen as words that describe an action in a scene. In the case of the sentence "*"In the center of the valley, an ancient, yet magnificent, tree grows on the bank of a modest river."*, the system must interpret the verb grows as referring to the tree. It refers to the action of the tree, which unmoved by its old age, is still growing. An increasingly popular NLP approach to understanding such relationships is called Dependency Parsing, and it builds upon the notion of dependency grammar.
Modern dependency grammar frameworks build upon the works of the French linguist Lucien Tesnière. Tesnière (1959) proposed the notion of verb valency or valence. Verb valency is the dependencies between, what Lucien referred to as dependents of the verb, and the verb itself. A sentence in that regard, is constructed from a verb, with a number of dependents. Tesnière

(1959) argued that we, as humans, create connections between words, and from the sum of these connections form the structure of a sentence. This structure establishes what is referred to as *dependencies* between words. Each *dependency* connections links an inferior term (the dependent) with a superior term (the head). In figure 3.2 a multitude of such relationships can be seen. Between the head and its dependents, in a dependency, there exists a binary asymmetric relationship. In Computational Linguistics, each arrow is connected to a specific dependency type. In figure 3.2 the arrow between the verb *is* and the noun *Tree* is a subject dependency, while the arrow between the preposition *near* and the noun *river*, is a postnominal (a modifier or argument). Interesting to note is that it is common practice to insert an artificial word, '*ROOT*', before the first word in a sentence, and make the verb of said sentence be dependent on the root. The reason for this, is to simplify the formal definition and the technical implementation [26].

Figure 3.2: Dependency Graph
Each arrow is pointing from head to the dependent

In dependency parsing, there are broadly speaking two main approaches which can be seen as a data-driven approach and a grammar-based approach. The data-driven approach relies heavily on machine learning, as is trained on linguistic data for the purpose of parsing sentences. The grammar-based approach relies on a formal grammar, that defines a formal language, with the purpose of determining whether or not a given input sentence is part of the defined language [26]. The data-driven approach utilizes supervised machine learning, where annotated linguistic data is required for training. This means that we need to obtain a lexical resource with such annotated data, for the Danish language, as it seems beyond the scope of this project, and our capabilities, to either define a formal language or annotate linguistic data ourselves.

As is evident from our research into the field of Natural Language Processing, it is an extensive field and there are many steps that needs be taken to fully realize *The Wor(l)d Builders* application. Therefore, we will conduct an iterative Design Process in which we seek to uncover the best solution to test the prospects of this thesis.

3.4 Design Requirements

To conclude on the research chapter, we present a list of summarized requirements based on the key points presented in the Instructional Technique Key Points and Didactic Perspective sections. These are the requirements that we deem to be the most essential for the first iteration of the design of a prototype application.

- The application must be based on process writing (pre-writing, writing and post-writing).
- The pupils must become actors, with meaningful choices within the application.
- The application must, through formative evaluation, induce self-evaluation and reflection in the pupil.
- The application must include a narrative that provides context for the written assignment. Furthermore, the narrative must be constructed such that the didactic strategies and writing purpose can be easily embedded.
- The application should encourages social interaction.
- Graphics and GUI should be simple, as to not distract from the writing assignment.
- The application should be centered around exploration and problem solving.
- For Natural Language Processing, different lexical resources are required:
  - To train a Word2Vec model: a corpus of text, fitting the context of *The Wor(l)d Builders*. E.g. certain Wikipedia pages, and fairy tales.
  - To train a Dependency Parser: Annotated Linguistics data, for the Danish language.
- Required Natural Language Processing Methods:
  - Part-of-Speech Tagging
  - Word Lemmas
  - Word Stemming
  - Word and Sentence Tokenization

We propose a design for a game based educational writing application, in which pupils will be encouraged, through visual feedback, to write sentences and texts. The narrative and mechanics of the application will be structured around *process writing*, and will provide context and writing purpose for the written assignment. The pupils will write input sentences, that will be analyzed through natural language processing which in turn will enable the game to provide visual feedback. The feedback will be of a formative nature, as pupils will be able to see the consequence of their writings (making them actors) through said visualization. A visualization which purpose is to foster reflections. As a pupil is producing a sentence, the subject of said sentence will ideally have some level of description. A pupil might be writing about a tree that grows in a park. The manner in which the subject of the sentence (i.e. the tree) is described will have a direct influence on how it is presented, visually, in the game. Based on the pupil's choice of words and sentence composition, the aesthetics of the scene will change. If the pupil describes the tree as beautiful and green, then the tree will appear as such, while a lack of description will present the tree as bland. Once a pupil has completed a writing segment (be it sentence or text), it is the "visual story" that unfolds before them, that encourage reflections. Writing better and more coherent segments, will then improve or alter the visual outcome, thus providing them with something tangible to foster an understanding of the importance of how they communicate through written language. This understanding should encourage pupils to improve upon their writing capabilities, through self-evaluation of the formative evaluation and reflections on the visual outcome. Furthermore, as each pupil writes their own visual story the game will encourage social interaction between classmates, as they compare, explain, and become inspired by each others choice of words and visual outcome.

This thesis will not be a direct attempt to solve the issue of motivating pupils to write more, but should be seen as the first step in a Design that combines the different methods we have uncovered in the research analysis. The design will be implemented in *The Wor(l)d Builders*, and tested in accordance with the Methods chapter (see Methods chapter). Specifically, we raise the following final problem statement:

To what extent can The Wor(l)d Builders application be taught to interpret sentences, written in Danish, and visualize the objects being described in said sentence

Elements of the educational writing game, such as modeling techniques, graphics and narration, will be discussed in the design chapter.

Chapter 4

Design

The design requirements (see Design Requirements section) will be used to design a prototype that can answer the final problem statement. This will be done through integrating didactic strategies paired with instructional techniques, as well as a thorough implementation of a natural language processing system. This chapter will first present our design of the didactic and game based elements, and then proceed to an explanation of our choices within natural language processing. Furthermore, we hope to create collaboration and social interaction between classmates when using the application without directly stating the fact in the application itself.

4.1 Game Based Learning Approach

In the Writing & Didactics section we found that process writing is the optimal strategy to teach writing. Process writing allows formative evaluation, that can create self-evaluation through constant feedback throughout the writing process. An application is the obvious choice to mediate process writing, with its ability to provide immediate feedback to the individual pupil. Therefore, the application will be structured around process writing and will include:

- A pre-writing stage for gathering information
- A writing stage with visual feedback. The stage will serve as the primary function of the application
- A post-writing stage where the pupil will be encouraged to reflect upon their writing.

Fig. 4.1 presents the process that a pupil will go through in the application. It is based on the works of [12]. Each step relates to a step in process writing, and can be used as a reference for the rest of this chapter. The

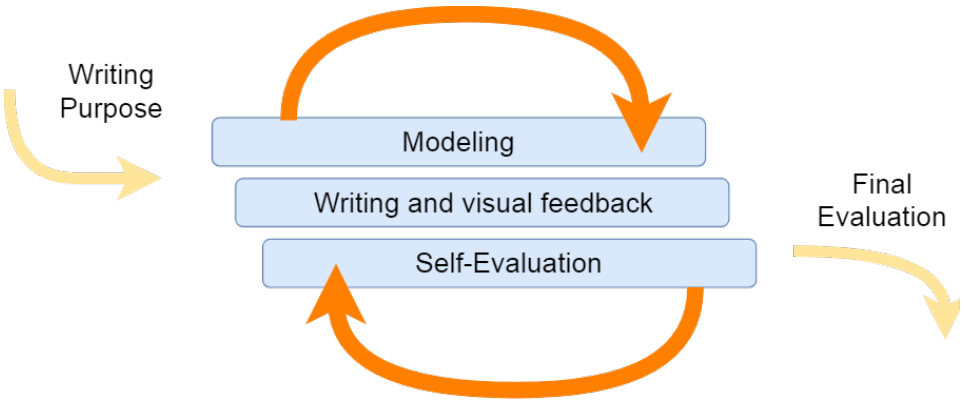

Figure 4.1: The Writing Cycle

left most arrow is *step 1*, and represents the writing purpose. *Step 2, 3* and *4*, are present in the middle of the figure. These are modeling, writing and visual feedback, and self-evaluation. These steps are represented by two arrows, enveloping their respective boxes. The arrows form a circle, and represent the iterative nature of process writing, where revision, feedback, and self-evaluation, will lead to more writing. The right arrow is *step 5*, and represents the final evaluation that happens at the end of the post-writing stage, when the assignment is deemed to be finished. The different stages will be elaborated on in the subsections below.

In accordance with the research by [24], both boys and girls enjoy exploration, and to some extent, problem solving as a game mechanic (Instructional Techniques). For the purpose of this thesis we have chosen a space setting, where pupils can explore a galaxy and draw inspiration from other planets to solve the problem of bringing life to their own planet.

4.1.1 Narrative

The context in which writing occurs is a prominent aspect of providing purpose to an assignment. As argued by Brok et al. (2015), the writing purpose is one of the most essential aspects of motivating ones pupils, as it ascends a task from mere homework to something meaningful. Providing pupils with a context for their assignment increases the likelihood of them engaging in the task, and opens up for the opportunity to use modeling techniques for inspiration. Besides the writing purpose, an interesting narrative in itself can also increase motivation, as stated by Wouters and van Oostendorp (2017). The most important aspect of the narrative in a game based application is content integration, where the writing purpose is made part of the setting. Embedding this purpose into the narrative and setting should greatly enhance the encouragement of the writing task, where this task will be the

written gameplay. The application will introduce a simple narrative, that will engage the pupil in a story of survival. *An alien race has been living in a large spaceship for thousands of years, but now seek to reclaim their old home planet. The problem is that the planet is now desolate, and by no means fit to sustain life.* Making the planet able to sustain life again introduces the writing purpose, as pupils can now directly influence the planet through their writing. *The aliens decide to rebuild the planets environment and nature, but they have unfortunately forgotten how these function. The pupil, in control of the main computer and navigation system of the spaceship, must then travel through the stars in search of other planets, that can inspire and teach them how to create their own planet.* These other planets serve as the aforementioned modeling techniques. The home planet will be rebuild through written input in the main computer's *creation text prompt*. The narrative will serve as the foundation for the learning outcome, providing both purpose and context to the pupil's writings. The pupils themselves will be writing the rest of the narrative, through their own short stories about the planet. Their own writings will create a semi-emergent narrative where pupils will feel a sense of agency in their ability to affect the world. An agency which should also serve as a motivational factor.

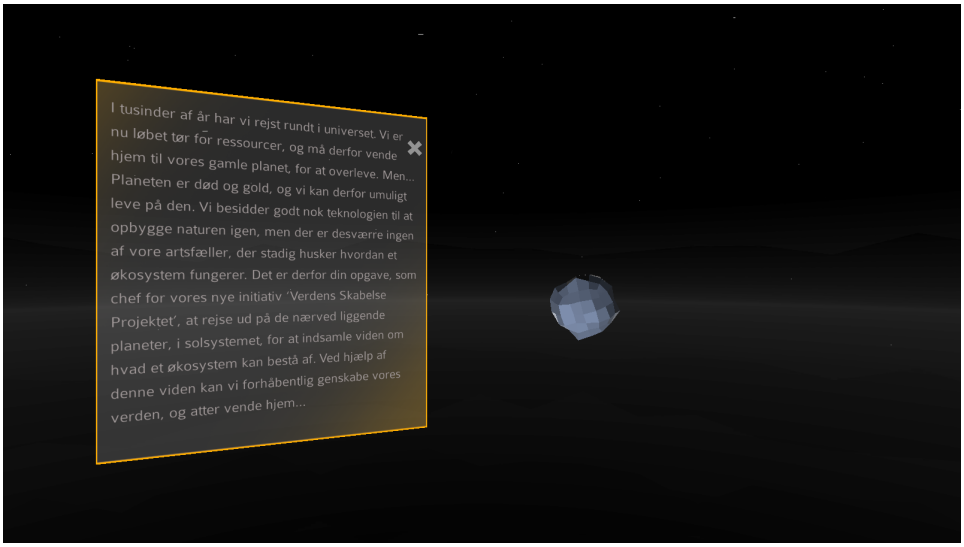

Figure 4.2: Screen displaying intro to narrative and objective

4.1.2 Pre-writing

The purpose of the pre-writing stage is to provide pupils with a description, and more importantly, a purpose for their writing. A clear and inspiring purpose should foster reflections on the process and allow pupils to properly

plan their task. The pupil will be presented with the main goal of creating life on a planet. It is important that this goal clearly relates to the principles presented by [12]. As stated in the Writing & Didactics the principles are used to provide context and purpose for the goal of an assignment. These principles will be integrated into the context of the application's simple narrative, and will help with creating motivation through a clear purpose.

Writing purpose Relates to the creation of a pupils personal motivation to do the assignment. The purpose is survival of your planet, and thus the text should be written in order for your planet to come to life. As you further elaborate your text, the 'quality' and nuances of the planet will increase, as explained in Writing & Didactics.

Educational purpose Relates to the learning objective. In the case of *The Wor(l)d Builders* the objective is to encourage pupils to improve their writing capabilities. When writing for their planet, their choice of words, sentence composition, and content of the text all build towards a more complete and unique planet. Through this notion, pupils will learn how to improve upon these when creating their planet, and be encouraged to do so through the visual feedback.

Receiver Asks the question of who the writing is meant for. For pupils, the extrinsic receiver is related to the writing purpose and narrative, where the planet and alien race is the receiver. Intrinsic motivation for the pupil could see themselves as the receiver of their own written improvements, as they experience how their writings can make a planet flourish. The latter is, as mentioned, what we hope to achieve with *The Wor(l)d Builders*.

Content Relates to the content of the task. In *The Wor(l)d Builders* we use visual feedback to foster reflections in the pupil, as these have a direct link to motivation and learning Writing & Didactics. Therefore, the pupil's written texts will be descriptive, using adjectives to explain the aesthetics of the different elements on their planet. Furthermore, the structure of the text will matter since the text will be written as a short story, where the order of events are important (e.g. a planet needing water first, in order for plants and trees to grow).

Action Relates to the completion of the task itself, which in the case of *The Wor(l)d Builders*, can be see as the gameplay and interaction with the application. Pupils will be writing short stories about their planets, as the planet comes to life before them. Much of the interactions will then include revision and rewriting of the text to improve upon their visual outcome. Revision and rewriting is a natural part of formative evaluation due to reflections, and is one of the important aspects of

process writing Writing & Didactics. Pupils will compare their planets and texts with each other, and be inspired in their own creation through this interaction. In the future, such an application could be used to create a mini solar system in a classroom, where each pupil has their own planet.

As stated above, the principles will be integrated into the narrative, and will appear under the writing purpose step in pre-writing, as it appears on fig. 4.1. After displaying the writing purpose (creation of their planet), the application will proceed to elaborate on how this feat is to be accomplished, through prompting the pupil to explore other planets that are already developed. As stated in Writing & Didactics, it is important that pupils receive examples that can help and inspire them to create their own writing. These modeling techniques are essential for a writing assignment, and are easy to integrate into the narrative and interactions of the *The Wor(l)d Builders* application. Each planet will include elements such as trees, water and mountains. The pupil can investigate these elements. To investigate, the pupil can *scan* an element (see fig. 4.3), which will prompt a text describing the element in question (see Appendix E). The text are written to reflect the elements role in an ecosystem. This will serve as the general modeling technique for all interactable elements on each planet. The visualization of the elements are meant to contextualize the text prompts, and hereby yield a higher level of understanding of the explanations. Paired with the modeling text, these elements should serve as inspiration as to how pupils could write about their own planets. Planning is an important aspect of pre-writing as well, and will be done through a notebook, which pupils can use to note their observations of the planets they explore, thus collecting data to plan their own planet. The notebook also encourages pupils to take advantage of the modeling techniques they are presented with. Once the pupil feels ready he/she can proceed to their own planet and engage in their writing assignment. As the modeling techniques should be a constant presence to aid and inspire, the pupil can go back and explore planets whenever they wish to during their writing process.

4.1.3 Writing

In Design Requirements, a prominent aspect of the application is the visual feedback (the visual representation of your text, namely the different elements on your planet), which will provide the pupils with a purpose for their text, and a medium to encourage them to improve their writing capabilities. The writing and post-writing stages are represented by *step 2, 3,* and *4* in fig. 4.1. As mentioned by [12], the goal should always be present such that pupils are never in doubt of the assignment. We will implement a button, that when pressed displays the main goal of the written assignment.

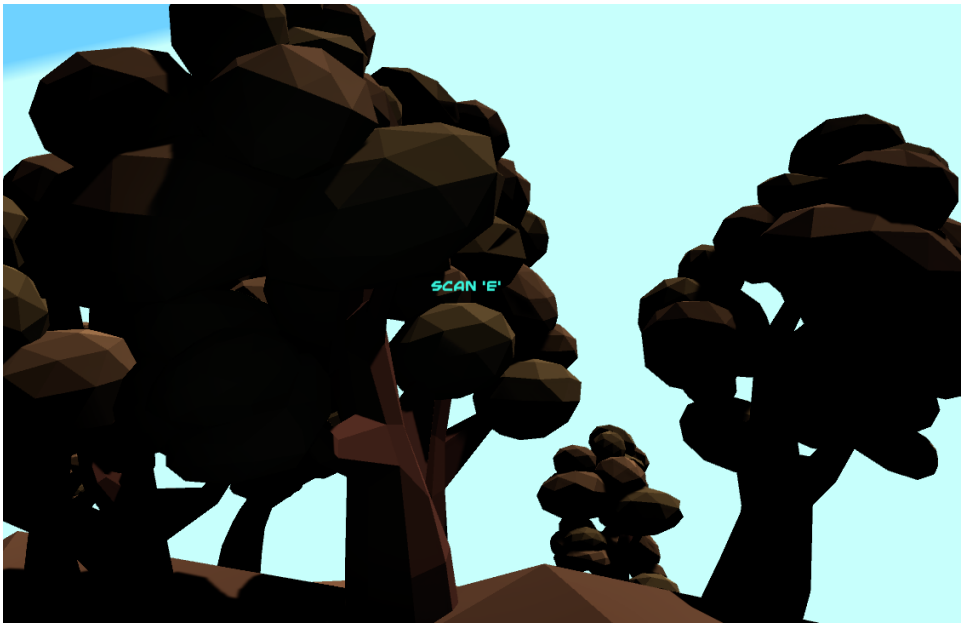

Figure 4.3: Scanning an object

It can be pressed whenever during the writing process. The button will also display a reminder that one can always go back and explore planets. The text interactions with visual feedback will be described more thoroughly in the natural language processing section.

4.1.4 Post-writing

The purpose of post-writing is to finish and evaluate a written assignment, but in process writing the evaluation takes place constantly throughout the process (See Pre-writing). The post-writing stage is represented by *step 4* and *step 5* in fig. 4.1 With formative evaluation, the constant feedback should foster reflections through self-evaluation. In process writing the teacher serves as a constant guide to help pupils with their self-evaluation, making them an essential part of the post-writing stage as well. Peer feedback is encouraged as a great method for inspiration and motivation, especially for pupils that are less fond of writing. Furthermore, revision and rewriting is essential to process writing as an iterative process, and is the product of self-evaluation and teacher/peer feedback. The post-writing stage of the application will revolve around a three step cycle, that consists of self-evaluation, peer and teacher feedback, and revision/re-writing. Self-evaluation will be the personal immediate feedback from the visual representation, where the pupils themselves will rewrite their text based on their

own reflections of their creation. Peer feedback is not mandatory, but is encouraged as a means to foster reflections through comparisons and inspiration. Finally the pupil will be encouraged to rewrite their texts. Rewriting should occur automatically as they revise their work. Rewriting is as simple as deleting words, sentences or segments of your text and rewrite, just as you would in a program such as Microsoft Word. This accessibility should encourage multiple iterations of revision and rewriting through their self-evaluation. *Step 5* (see fig. 4.1), *the final evaluation*, occurs once the pupil is satisfied with their planet, and has successfully made the planet able to sustain life. The writings of the pupils and the final visuals of their planet, should be accessible to their teacher at this stage, so that the teacher may partake in the evaluation.

4.1.5 Graphics

The graphics are important for the visual feedback to become a proper representation of the pupil's written text. Realistic and beautiful graphics would arguably create the optimal visual experience, which could help encourage pupils in their writing. Unfortunately, creating and customizing such graphics can be a difficult process and is beyond the scope of this project. Such graphics could be implemented in a future iteration of the application. In the Instructional Techniques [51] argue that simple, cartoon like graphics are best suited for educational games. They provide a better learning experience than realistic graphics, as the latter can be a distraction. As such, while the use of low polygon and cartoon like graphics might not produce the most aesthetically pleasing outcome, they may be sufficient for the purpose of *The Wor(l)d Builders* prototype, and provide a good experience of the learning environment.
When a pupil submits a sentence or text to the applications text prompt, the elements on the planet should appear as though they are growing. If implemented properly it will enhance the effect of the pupils creating their own world. For the purpose of the prototype of this thesis, we will be utilizing pre-made 3D assets, as we do not have any 3D artists at hand.

4.2 Game Engine

The Unity3d game engine will be used for the *game* aspect of *The Wor(l)d Builders* application. While other game engines such as Unreal and CryEngine can provide a better framework for creating visually astounding game worlds, the limitation of the assets creation for this project will not allow for such graphics, and thusly the Unity3d engine suits the purpose quite well. The Game Engine is needed for the specific purpose of:

User Input Unity has a neat User Interface script called *Input Field*

which takes input from keyboards, and can trigger other functions upon completion of e.g. sentence writing.

Visual Representation Generating objects based on User Input.

Modeling Displaying a variety of text based prompts to help pupils.

Context Presenting a narrative, and emphasizing causality of the pupils writings.

4.2.1 Planet Design

The explorable planets were designed with simplicity in mind, using low polygon graphics (see section Graphics). The planets needed to include all the elements, that a pupil can spawn on their own planet, with text prompts describing each element once interacted with. We used a variety of assets from the Unity Asset Store [42][45][16] as well as self-made 3d low poly planets. The asset packs comes with a variety of 3D models to choose from, which allowed us to create three planets with distinct themes:

Forest Planet A green planet, with oak and poplar trees. The planet also has rocks and a small water stream (see fig. 4.4).

Desert Planet A yellow/orange coloured, and mostly empty, planet. At the peak of the planet is a small oasis of palm trees, rocks and a small water hole (see fig. 4.5).

Snow Planet A white/light blue planet, with many snow covered pine trees. The planet also has different mountains, bushes, frozen lakes and glaciers (see fig. 4.6).

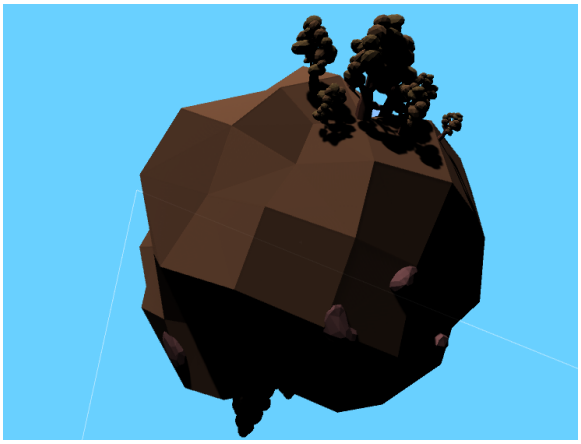

Figure 4.4: The Forest Planet

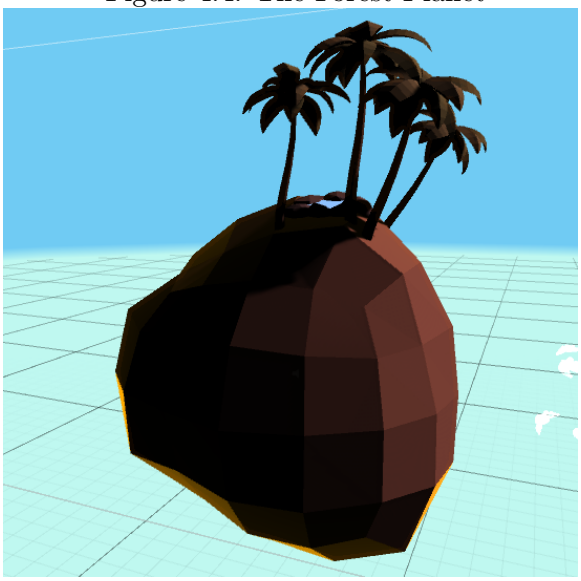

Figure 4.5: The Desert Planet

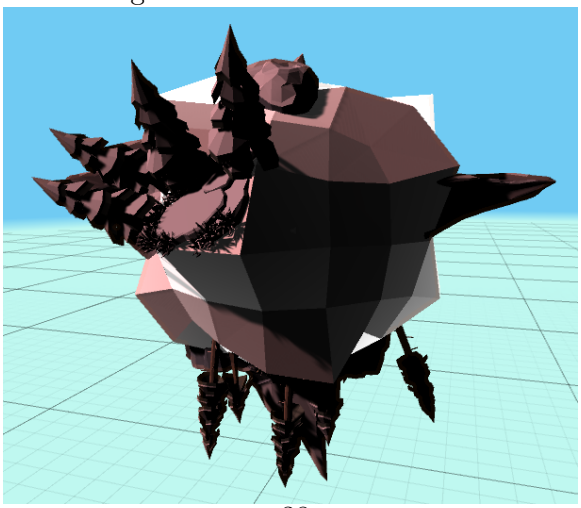

Figure 4.6: The Snow Planet

4.2.2 Graphical User Interface and Audio

As per the design requirements (see Design Requirements section), the graphical user interface (GUI) had to be simple and unintrusive. We used a premade GUI from ([42]) for the start menu (fig. 6.1, and buttons, and created the input field archive function ourselves. The archive function is a button, that when pressed displays the pupils previous sentences (fig. 4.7). The function can be used to edit these sentences, and serves as a practical method of finding inspiration in previous work. The application included a small amount of audio, namely ambience soundscapes for each planet, feedback sounds (when clicking GUI buttons), and footstep sounds when walking around on a planet. Sound assets acquired through Unity Asset Store [21][9], and the *Hollywood Edge Premiere* sound library. Ambiences were edited and mixed by us, to fit with the setting.

Figure 4.7: The home planet GUI with the archive open

4.3 Natural Language Processing

As per the Design Requirements established through analysis, a variety of Natural Language Processing (NLP) methods were used in the design process of the application (see section Natural Language Processing). To implement the requirements we used Python 3.5 because it is the most commonly used Programming Language in Machine Learning and Natural Language Processing. Therefore, Python has a variety of freely available tools, such as the Natural Language Tool Kit (NLTK) and Polyglot. The Python side of the application needed to be able to handle:

Tokenization Both the tokenization of text to sentences as well as sentences to words.

Stemming Words stems, and lemmas are quite useful to limit the vocabulary of the application, to only use the root of words.

Part-of-Speech The system is required to know which words are nouns, adjectives and numerals etc.

Word2Vec The representation of words as vectors, enabling semantic meaning extraction.

Dependency Parsing Understanding the relationship between words in a given sentence.

4.3.1 The Design Process of Natural Language Processing

In our search for information about Natural Language Processing (NLP) we also needed to get a better understanding of the concepts, and the field itself in order to develop an NLP application, and understand the various outputs of the Natural Language Processing Tool Kit(NLTK) and Polyglot. To do this, we used a variety of resources available via the internet. First and foremost we used Stanford's NLP Course which is available on *www.youtube.com*.
Following these courses we started to develop a Word2Vec representation of Danish words. The corpora was collected from Wikipedia and Gutenberg (approximately *500.000* sentences). There were, however, some issues with retrieving data directly from Wikipedia and Gutenberg. Wikipedia include a lot of wiki specific symbols that does not carry any linguistics meaning, and needs to be filtered out. Gutenberg is a database of texts that are at least sixty years old, meaning that the language used in those texts does not resemble what is used in modern Danish. This included the manner in which sentences were structured and the representation of the letters *æ, ø* and *å*, which were written as *ae, oe* and *aa*. These letters were fairly simple to change through the use of regular expressions, replacing them with modern Danish letters. After doing so, we needed to prepare the words for training. The first step is *Tokenization* and *Stemming*. The Natural Language Processing Toolkit (nltk), provided us with a tool for word and sentence tokenization. Some of its limitations is however, that in the Danish word *Hr.* (i.e. Mr.) the sentence tokenizer would determine that this was the end of a given sentence, which in most cases is incorrect since the abbreviation is usually used in advance of a surname. Furthermore, the nltk stemmer, '*snowball*', has somewhat limited applicability and failed to recognizes many suffixes apart from 'e', yielding too many entries of the same word in the data set. Too many entries will, as we learned, ultimately affect the precision of the Word2Vec model. The fewer contexts a given word is '*seen*' in,

will affect the features of the word, hereby loosing semantic meaning.
Training a Word2Vec model on the corpora collected from Wikipedia and Gutenberg , yielded unsatisfactory results regarding word similarities. The purpose of these similarities was to check if the words in an input sentence shared semantic meaning with the words we had predefined, describing the objects in a scene (e.g. *Beautiful, Large* and *Small*). This way, we meant to enable users (i.e. pupils) to write more freely, as they would not have to write the specific predefined adjectives to describe something as being e.g. *Beautiful*. From these results, and after testing the application on various words, we realized that the danish corpora that is available to the general public, and students on Aalborg University, is not enough to train models for the purpose of this project. Furthermore, we found that the semantic features of words in a Word2Vec model might be to abstract to convert to in Unity. This is because Word2Vec is generally used to describe some semantic relationship between a word, given another set of words. If a pupil write a sentence, the application must be able to understand said sentence, and not an abstraction of the meaning of each word in the sentence. The concreteness of the words is quite important (disregarding the lemma of a word), as we aim to design the application towards understanding a description of concrete objects, and not the semantic relatedness of words in a sentences, given another set of words. This is why we deem a Word2Vec model less applicable to the purpose of *The Wor(l)d Builders*.

After having learned that Dependency Parsing would require access to labelled data, that linguists have to manually create, we tried to find a lexical resource which a model could be trained on. We meant to train the Standford MaltParser, but did not succeed in doing so as we were unable to obtain labelled data. Manning and Schütze (1999) and Kübler et al. (2009) stress the need for experience within the fields of computational linguistics, and linguistics to have the proper foundation for starting to learn about NLP. This became evident during this design process, as we implemented a variety of NLP methods. This is why we ask Prof. Gitte Rasmussen, *head of Center for Social Practices and Cognition, University of Southern Denmark (SDU)*, who set us in contact with Research Assoc. Prof. Eckhard Bick, *Department of Language and Communciaton, SDU.* Through Research Assoc. Prof Eckhard Bick we were able to obtain access to the VISL, which is part of the Visual Interactive Syntax Learning (VISL) tool. VISL started as a learning initiative in 1996, and has since been funded by various contributors who have made it possible for the VISL team, to expand and implement a variety of Natural Language Processing tools, including, but not limited to the Dependency parsing. With access to the VISL API, we are able to derive lemmas from words, determine part-of-speech and the dependencies that exist between the words in a given input sentence. This is why we chose to solely use the VISL API for Natural Language Processing We do

however, need to create a data analysis script to process the output from the VISL parser, in order to turn it in to command keywords that unity can interpret. The implementation of this is described in the Implementation Chapter.

Chapter 5

Methods

This chapter will describe how we intend to design the experiments meant to answer our problem statements:

To what extent can The Wor(l)d Builders application be taught to interpret sentences, written in Danish, and visualize the objects being described in said sentence
Specifically, we seek to reject the following hypothesis:

H0 *The Wor(l)d Builders* application cannot visualize objects of sentences, describing said objects, within the context of the applications narrative.

5.1 Measurement

Seeing as, to the best of our knowledge, no one has previously created a design similar to that of *The Wor(l)d Builders*, we propose an accuracy measure to answer the problem statement. Our intention with this measure is to get insight into what we might improve upon in the design of the application, and to see whether the implemented (see Implementation chapter) methods are applicable in The Wor(l)d Builders. By logging the input sentences, and output strings (i.e. processed sentence), we seek to measure the accuracy of the application by calculating:

$$Precision = \frac{tp}{tp + fp}$$

$$Recall = \frac{tp}{tp + fn}$$

$$True\ Negative\ Rate = \frac{tn}{tn + fp}$$

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn}$$

The criteria for the above measurements are as follows:

True Positive A sentence being processed, and visualized correctly.

False Positive A sentence that has been accepted even though it contains an object, that is *not* part of the applications object vocabulary (i.e. each object is tied to a noun).

False Negative A sentence that has been denied, but the object in question *is* part of the applications object vocabulary.

True Negative Gibberish sentences and spelling errors.

We expect the accuracy measure to be heavily influenced by false positives. This is because we do not seek to filter out nouns and adjectives, in the NLP side of the application, that does not have an impact on the visual outcome. Therefore, we expect test participants to challenge the system and go beyond the predefined nouns and adjectives, that can be be visualized in the application. The accuracy measure is meant as a catalyst for discussion, and can reflect how well we adapted the '*game*' aspect of the application, to the *Natural Language Processing* aspect. This adaptation, is consequently quite influential on the performance of the system.

The long term purpose of the application is to create an educational game, that encourages pupils to improve their writing capabilities. Based on the current design, we argue that the prototype will be an application inspired by game based learning, but lacks the ludic element that usually make a game. While not the main purpose of this thesis and the experiment, we will look for evidence concerning the utility of the finished application. It would be interesting to know whether the visual feedback created by the application, could exceed regular classroom methods regarding the encouragement of pupils to improve their writing capabilities. To test whether the application can provide a visualization of sentences, and whether we can reject the H0, a mixed quantitative and qualitative experiment will be conducted. A group of participants will be tasked with using the application, for a maximum of 20 minutes. We believe that at the current state of the prototype, an average play time of 10-15 minutes can be expected, thus we will set the limit at 20 minutes. During the test session, the application will log all of the participants' written input, as mentioned above. Subsequently, we will give each participant a questionnaire, which concerns their experience in the application. The questionnaire will include likert items with a range of 1-7 ("completely disagree" to "completely agree"). The 1-7 scale is chosen in order to increase accuracy through higher variance, and because we want to give participants the option to choose a neutral middle value in occurrences where the individual feels undecided or divided. Except for two single answer questions (yes/no), each item will be followed by a qualitative

question, that encourages participants to elaborate on their reply (see Appendix A B for the questionnaires). The questions will be based on The Writing Cycle (see fig. 4.1). As mentioned, the model serves as the basis of our design, and presents the cycle, that a participant will go through in the application. The questions concern:

Writing Purpose The purpose of the application.

Writing Visualization of objects.

Modeling Do you explore the game world, to understand the concepts, and use those to complete the assignment.

Self-evaluation Do they revisit content, and/or reflect upon the visualization of the objects on their planet, to improve their perceived quality of said planet, by rewriting previously written sentences.

The responses will be analyzed using meaning condensation. Through this meaning condensation we will derive at a set of themes, which will be used to determine the participants general thoughts regarding the different steps of the Writing Cycle.

5.2 Experiment

This section will present how the experiment was carried out. The test was conducted over the course of four work days. Two days were spent testing at the *NEXT - Uddannelse København* highschool in Ishøj (n=29). One day was spent testing on middle school pupils, who had been invited to participate in experiments at Aalborg University Copenhagen (n=14). The fourth test day was conducted in Blommenslyst, Lille Uberud, on participants from both highschool (n=2) and middle school (n=6). We brought two desktop computers, and invited participants to test, two at a time. We created a protocol that we followed during each test, providing test participants with as similar an experience as possible. Participants were introduced to the controls and the task at hand, and were told that they were always allowed to asks questions. After the brief introduction, participants were told to press the 'Intro' button in the start menu, which would lead them to a text box explaining the narrative (extrinsic writing purpose). Once done with the reading, we gave participants a short introduction to the UI buttons, their home planet and the planets they could travel to. Furthermore, they could always refer to an instructional paper that we had put next to each computer. The instructional paper explained controls, and included a small guide to the adjectives and numerals available in the application. Finally the instructional paper included an example of an input sentence, that participants could refer to if they were in doubt. The instructional paper can

be found in Appendix C. Participants were told that they were welcome to finish the test once they felt that they were done with creating their planet. After they were finished with the application, we pointed them towards two laptop computers for the questionnaires. It should be noted that we removed the qualitative items in the questionnaire for the middle school participants. We did not want to stress them with writing long replies, after having just written in the application. As we have learned in the Introduction section, the meta cognitive skills of children this age are not fully developed. We believed that they would not be able to properly reflect on their experience in the application, through a qualitative response.

Chapter 6

Implementation

In this chapter, we will present an explanation of the system in its entirety, as well as delve into the different subsystem that was used for testing in this Master Thesis.

6.1 Unity3D - *The Wor(l)d Builders*

6.1.0.1 Unity Engine & Python

When deciding to work with Python and Unity we knew there would be some compatibility issues due to the Unity3D Engine not natively supporting Python. This is why we used a local TCP client-to-server based connection to send data back and forth between C# scripts in Unity, and a standalone Python build. This solution had the potential of limiting the feasibility of testing because the computer on which the application was to be run, had to have Python 3.5 installed. To make testing more accessible we ended up using *PyInstaller* to build an executable of the Python code, which enabled usage of the code on computers without having to install Python.

6.1.1 Graphical User Interface

The first thing the test participants were exposed to was a Start Menu (see fig. 6.1), from where they could start or exit the game, and get an introduction to the narrative, while at the same time being introduced to their assignment, in *The Wor(l)d Builders*. The game world was divided into four scenes:

- Start Menu
- Planetary Overview
  - Here they could click a planet to visit, or go to their home planet.

- Planet Exploration
- Home Planet

It was in the fourth scene they were able to write sentences, which would generate objects on their planet. All the written sentences (of a length longer than 8 characters, a minimum we set to avoid empty strings and single words) was collected in a '*folder*' which were accessible through a folder icon when the fourth scene was active. We did not implement the button mentioned in the Writing section in the Design chapter, but instead included buttons that the player could use to return to the start menu and the planetary overview.

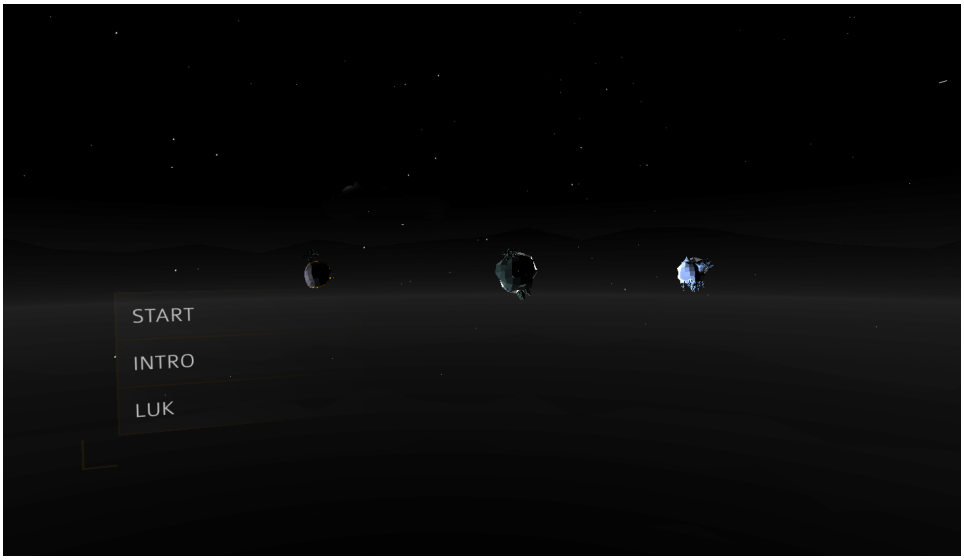

Figure 6.1: The Starting Menu

6.1.2 Manipulating Game Objects

While the actual text processing was done through VISL, and Python, Unity needed methods to handle the data output from Python. To do this we wrote a scriptable object class, a class that were attached to the different objects that was to be instantiated, in the game world, through user written sentences. This scriptable object had a range of modifiers that would alter the object, which were the subject of an input sentence. One would scale the object, another would change the colour of the mesh, and a third would change the mesh of the model. These modifiers were dependent on two things: 1. The noun, and 2. the adjectives that were written in the same sentence. One of each modifier were attached to the *spawnable* gameobjects

and a noun was manually written in the inspector to each of those modifiers. When ever a string was received from Python, the code would look for objects associated with the given subject of the sentence and modified according to the adjectives that were also present in said sentence (e.g The word *Lille* (i.e. *small*) would scale the appropriate objects).

6.1.3 Objects and Interactables

As can be seen in the Planet Design section each planet was created with interactable objects refering each object, that a participant could create on their own planet. We decided to remove mountains, hills and bushes from the list of interactables. Altering these through adjectives, as mentioned above in the Manipulating Game Objects section, did not feel natural due to the design of the 3D models. Furthermore, spawning mountains and hills on the planet did not work as intended, due to how objects are spawned randomly on the vertices of the home planet mesh. How water is spawned on the home planet was altered due to the water asset we had access to, being a simple material that was added to a mesh. Attempting to spawn water on the home planet would then lead to one or more faces on the planet mesh simply turning blue. We instead opted to symbolize water by spawning a range of clouds around the planet, whenever a participant would attempt to create water. We derived at a final list of spawnable objects:

Trees A range of trees, including oak, poplar, palm, pine, and snow covered pine trees, which could be modified by adjectives.

Rocks A 3D model of a rock, which could be modified by adjectives.

Water Symbolized by a range of clouds surrounding the planet

Trees and rocks could then be modified as stated in the Manipulating Game Objects section above. The available adjectives were *little, small, big, large, green, blue,* and *red* (translated from danish). Furthermore, participants could decide the amount of said object, ranging from one, to one hundred.

Figure 6.2: An example of a home planet. Screenshot from test

Fig. 6.2. displays an example of a home planet with different objects added. The participant has decided to add red palm trees, blue pine trees, green poplar trees, and what seems to be a a rock at the bottom. Furthermore, the participant has added water, which can be seen by the circle of clouds surrounding the planet.

In the Design chapter we mentioned the notebook feature, which purpose was to allow pupils to note their observations of the objects. Unfortunately, the notebook did not make it to the prototype, due to difficulties in the UI set-up. Furthermore, we did not implement *step 5*, as seen in fig. 4.1, as we felt that the scarce amount of objects were not enough to fully satisfy the fantasy of creating a planet able to sustain life. Thus we opted to tell participants that they could finish whenever they felt their planet was done, as mentioned in the Methods chapter.

6.1.4 User Input Handler

There were multiple levels of input handling of the users input. We decided to have the user control the game through 1. Point and Click, 2. First Person Controller (i.e. WASD and Mouse Control) 3. Writing sentences. The point and click script used ray casts with different masks, enabling clickable planets and Graphical User Interface (GUI). The first person controller was a bit

different from what is usually used in the Unity Engine, because it has a predefined gravitational system. We wanted the player to have the feeling of visiting little island like planets, where they could cross the entire globe within a minute or two. This of course meant that we could not use Unity's gravitational system, because it would drag the player downwards on the Y-axis, and we needed the gravitational pull to come from the planet that they were visiting.

6.2 VISL - Dependency Parser

To access VISL dependency parser we used a PHP scripts that handled Challenge - Request authentication. The output of the VISL parser includes; lemmas, word class and dependency relationship with another word (i.e. a tag, and a point from the words position, to its dependent). All of this information is sent back to Python, after processing an input sentence, written by a pupil.

6.3 Python - Sentences to Commands

The Python side of the application consisted of two parts, a server script and a script that handled the output of the VISL dependency parser.

```
# Find et subject some ikke er en person
if re.search(r'\B\@<*SUBJ\>* \#\w+', sentence) and not re.search(r'\B PERS', sentence)
    subject = str(re.findall(r'\B\@<*SUBJ\>* \#\w+', sentence))
# Hvis der er en subject component, brug da denne som subject
if re.search(r'\B\@<*SC\>* \#\w+', sentence) and not re.search(r'\B ADJ', sentence):
    subject = str(re.findall(r'\B\@<*SC\>* \#\w+', sentence))
# Hvis der gøres brug af akkusativ, som samtidig er et navneord, brug da dette som subj
if re.search(r'\B\@<*ACC\>* \#\w+', sentence) and re.search(r'N NEU', sentence):
    subject = str(re.findall(r'\B\@<*ACC\>* \#\w+', sentence))
# Find numerals i sætningen
if re.search('NUM', sentence):
    numdata.append(clean_string(str(re.findall(r'\B\[+\w.* *\]', sentence))))
    numPointer.append(int(float(clean_string(str(re.findall(r'\-\>\d', sentence))))))
# Find adjektiv i sætningen
if re.search('ADJ', sentence):
    adjdata.append(clean_string(str(re.findall(r'\B\[+\w.* *\]', sentence))))
```

Figure 6.3: Data Analysis Function from Python

In order to interpret the semantic analysis of the VISL dependency parser, we wrote a simple data analysis function in Python (see Figure 6.3). This analysis was used to segment the output into words that had specific

types of dependencies, namely: numeral (NUM), subject (SUBJ), and adjectives (ADJ). Furthermore, we created an array of dependency pointers ($X \rightarrow Y$) through the use of regular expressions. The lemma of each words were used, instead of the word that was written by the pupil, for better generalization (e.g. *små* and *lille* would both be passed as *lille*). The numeral string affected the number of objects being spawned in the game world. These strings were compared to keys in a dictionary that contained the corresponding integer as a value ranging from *one* to *nineteen* and *twenty* to *hundred*, the latter being only in tens (i.e. twenty-one was not a valid numeral), a limitation we decided upon because we did not deem it necessary, at this stage of the implementation, for the pupils to have full control over the number of objects that would be instantiated in the game world.

Seeing as there are a multitude of sentence structures, the subject of the sentence can change, depending on how it is being expressed in a given sentence. There are also cases were the subject dependency may not be the target gameobject to be spawned. For instance, if a pupil would write (*'I want twenty large, green Oaks on my planet'*) the subject, would be *I* and not oaks. This is why, the data analysis in python determined the *target gameobject* from the following requirements:

- Subject (*That were NOT a person*)
- Subject Components (*That were NOT adjectives*)
- Accusative (*That were a Noun*)
- Noun phrases

After the *target gameobject* has been determined, we look for a word with a numeral dependency to determine how many gameobjects should be spawned. Lastly we look for any words with an adjective dependency with the noun of the *target gameobject*.

The final output that was sent from python to Unity was of the format:

$$NOUN, NUM, ADJ_{x1}, ADJ_{x2}...ADJ_{xn}$$

which could easily be separated in C# using a string split function, with ',' as delimiter.

Chapter 7

Results

The following chapter will present the results from our experiment. The participants (n=51), included 31 highschool students and 20 middle school pupils. The chapter will first go through the logged files from the playthroughs and then proceed to look at the likert items from the questionnaire. Finally the qualitative responses from the questionnaire will be mentioned, but their primary use will be as a supplement to the discussion.

7.1 Data Logs

In the data logs we found the following results: *True Positives = 249, True Negatives = 39, False Positives = 108, False Negatives = 53.* The accuracy (see Methods) of the data logs were calculated using the following formula:

$$Precision = \frac{249}{249 + 108} = 0.731$$

$$Recall = \frac{249}{249 + 53} = 0.847$$

$$True\ negative\ rate = \frac{39}{39 + 108} = 0.265$$

$$Accuracy = \frac{249 + 39}{249 + 39 + 108 + 53} = 0.674$$

The accuracy result shows that the application successfully visualized the described objects in 67% of all sentences (n=494) written by the test participants. The data logs can be found on the ZIP-file included in the *Digital Eksamen* hand-in.

7.2 Questionnaire Scales

This section will present the results from the likert scale part of the questionnaire (see Appendix D for all SPSS output). The following tests were done on the likert items only, which means that the yes/no questions were removed from the samples. Furthermore, the negatively weighted items were reversed such that their scale matches the rest. A mean was calculated for scores between positive and negative items that ask the same question, which connects the two items into a single item. The total means for each item was calculated in Microsoft Excel, and can be seen in fig. 7.1 As can

Items	Means
Item01	5.294
Item02	4.275
Item03	4.49
Item04	5.98
Item05	4.088
Item06	5.098
Item07	5.275
Item08	5.333
Item09	5.372

Figure 7.1: Table displaying the total means for each item

be seen on the table, participants rated *item04* the highest, which is the total item of the two items concerning whether they used the planets for inspiration (Modeling). *Item05* was rated lowest. This is the total item of the two items concerning whether they felt like sharing their planets with classmates (socializing and evaluation).

Although not the main purpose of the research, it could be interesting to learn what differences there were between the highschool test group and the middle school test group. These results could provide us with different indications of the prototype, such as difficulty, comprehension of the content, and motivation to undergo the task. We are comparing each likert item between the groups, to see whether there is a significant difference between the two. We are aware that there is a difference in the number of participants in each group, and will thus conduct two tests. One on the total, but unequal, amount of participants, and the other where 11 participants of the highschool sample are removed, to achieve an equal number of n=20 in each group. A Shapiro-Wilk test was conducted on each item to learn whether the data was normally distributed. The results from the test are shown in fig. 7.2

The Shapiro-Wilk test, and the table shows that most of the results are not normally distributed, except for a few items. When deciding whether a

Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
Gym01	.948	20	.339
Gym02	.971	20	.771
Gym03	.941	20	.246
Gym04	.895	20	.033
Gym05	.928	20	.144
Gym06	.884	20	.021
Gym07	.903	20	.047
Gym08	.917	20	.085
Gym09	.893	20	.031
Skole01	.898	20	.038
Skole02	.882	20	.019
Skole03	.949	20	.351
Skole04	.815	20	.001
Skole05	.932	20	.169
Skole06	.890	20	.027
Skole07	.905	20	.052
Skole08	.903	20	.047
Skole09	.877	20	.016

Figure 7.2: Output from SPSS displaying test of normality

test should be parametric or non-parametric, results from both groups need to be normally distributed before a parametric test is the valid option. In the table above, only likert item three and five are normally distributed in both groups (Gym03 - Skole03, Gym05 - Skole05). First an independent t-test was conducted on item three and five. No significant difference was found between item three for highschool students ($M = 4.92$, $SE = .43$), and item three for middle school pupils ($M = 4.62$, $SE = .36$), $t(31) = .52$, $p = .61$. The same occurs in item five for highschool students ($M = 4.58$, $SE = .33$) and item five for middle school children ($M = 4.18$, $SE = .30$), $t(31) = .88$, $p = .38$. The non-parametric data was tested using a Mann-Whitney U test, which is the equivalent of an independent t-test for non-parametric data.

Figure 7.3 shows the results from the Mann-Whitney U test. We included item three and five in the test out of curiosity, even though they were deemed to be parametric. The results show no significant differences in items between the two groups, except for item six and item eight (Var06

Test Statistics[a]

	Var01	Var02	Var03	Var04	Var05	Var06	Var07	Var08	Var09
Mann-Whitney U	281.500	280.500	281.000	281.500	294.500	183.000	293.000	184.500	287.000
Wilcoxon W	491.500	490.500	777.000	491.500	790.500	393.000	503.000	394.500	497.000
Z	-.556	-.574	-.564	-.572	-.302	-2.509	-.337	-2.489	-.456
Asymp. Sig. (2-tailed)	.578	.566	.572	.567	.763	.012	.736	.013	.649

Figure 7.3: Output from SPSS, displaying results from the Mann-Whitney U test

and Var08). Item six states " *the visual representation of my text, could motivate me to improve upon my writing*". The result of item six shows that highschool students ($Mr = 30.10$) differed from middle school children ($Mr = 19.65$) in their replies to the question $U = 183$, $s = .012$, $z = -2.50$, $r = -.35$. The r value denotes a medium effect size. Item eight states " *I would receive a better understanding of a given subject, had I been taught the subject through the game*". The results of item eight shows that highschool students ($Mr = 30.05$) differed from middle school children ($Mr = 19.73$) in their replies to the question $U = 184.5$, $s = .013$, $z = -2.49$, $r = -.35$. The r value denotes a medium effect size. The significant difference weights positively towards the highschool students in both items, meaning their replies were of a higher average value. When comparing the total data from both groups, the results show that highschool students ($Mr = 230.24$) differed from middle school children ($Mr = 214.13$) in their total replies $U = 22253.5$, $s = .038$, $z = -2.08$, $r = -.097$. The r value denotes a small effect size. The results here shows that there exists a significant difference between the two groups when comparing their replies to each question in total, even if this difference does appear to be quite small. Again, the difference weights positively towards the highschool students, but from the individual item results above, we know that the difference primarily occurs in item six and eight. Nevertheless, the overall results point towards a tendency to rate items slightly higher among the highschool group.

For the test with an equal amount of participants, 11 participants were randomly removed from the highschool group, leaving both groups with 20 participants. Figure 7.4 shows the results from the Mann-Whitney U test with equal participants.

The results with equal participants is very similar to the results from the test with an unequal amount of participants. A significant difference can only be found in item six and item eight. The significance value is lower for the two items, which should indicate a larger difference in the two items with equal participants. Item six shows the effect size $r = -.495$, and item eight shows the effect size $r = -.426$. The r value of item six denotes a large

Test Statistics[a]

	Var01	Var02	Var03	Var04	Var05	Var06	Var07	Var08	Var09
Mann-Whitney U	163.000	134.500	196.000	146.000	197.000	87.500	197.500	103.000	197.000
Wilcoxon W	373.000	344.500	406.000	356.000	407.000	297.500	407.500	313.000	407.000
Z	-1.012	-1.788	-.110	-1.552	-.082	-3.131	-.069	-2.691	-.083
Asymp. Sig. (2-tailed)	.311	.074	.913	.121	.935	.002	.945	.007	.934
Exact Sig. [2*(1-tailed Sig.)]	.327[b]	.076[b]	.925[b]	.149[b]	.947[b]	.002[b]	.947[b]	.008[b]	.947[b]

Figure 7.4: Output from SPSS, displaying results from the Mann-Whitney U test with equal participants

effect size, while the r value of item eight lies somewhere in the middle of a medium and large effect size, learning towards a large effect. The results from the Mann-Whitney U test of total replies with equal participants show that highschool students ($Mr = 197.18$) differ from middle school pupils ($Mr = 163.82$) in their replies $U = 13197.5$, $s = .002$, $z = -3.07$, $r = -.162$. The r value denotes a small effect size, similar to the test of total replies with unequal participants. The effect size is larger than the previous one, but is still closer to a small effect size than a medium.

From the likert results, we can conclude that, with an equal and unequal amount of participants, there were differences in two items, and a small, but significant, difference in total scores. The results indicate that high-school participants were slightly more fond, and positive of the experience, especially with the questions stated in item six and eight.

7.3 Qualitative Responses

As mentioned in the Methods chapter, the questionnaire was based on the Writing Cycle Model (see fig. 4.1). The likert items provide a measurable indication as to how participants felt about the different aspects of the application. The related qualitative responses can then provide us with more elaborated feedback. The responses were analyzed using meaning condensation. This section will give an overview of how participants felt about the application, split into the different steps of the Writing Cycle. Common themes from their responses to each step, have been derived from the meaning condensation. All replies can be found in Appendix D.

7.3.1 Writing Purpose

Themes are: *Learning/researching other planets and creating your own. Building an ecosystem. Testing the game* Participants all agreed that there was a purpose with the task they were given in the application. The majority agree that the purpose was to build your own planet, using other

planets as inspiration. A few participants reported that the purpose was to learn about, and build, a functional ecosystem. Finally, a few participants believed the purpose to be testing the game.

7.3.2 Modeling

Themes are: *Planets used multiple times. Planets widely used as inspiration. Planets used for spelling correction. Planets were empty and did not have enough content.* The questions concerning the modeling aspect of the application yielded positive results, where the majority of participants agree that they used the other planets multiple times. The planets served as the main source of inspiration, as reported by the participants. Participants report using the same planets multiple times, to correct spelling mistakes and to explore what other elements the planet might offer. Few participants report that the planets were too empty, and that the given content was too limited.

7.3.3 Writing and visual feedback

Themes are: *Visual feedback and descriptions were limited. The small amount of visual feedback was enjoyable. Descriptions had an effect on the visual outcome.* Participants reported being somewhat excited about the possibility of the visual outcome, but were discouraged by the limited degree to which they could customize the objects (i.e. locations, colours etc.). They also thought that there were strict rules regarding formulation of sentences. Nevertheless, participants reported that their descriptions had an effect on the visual outcome, and that this effect was enjoyable.

7.3.4 Self-evaluation

Themes are: *Rewriting to correct spelling mistakes. Revision of previous sentences for inspiration. Revision and rewriting to improve descriptions. System was too limited to require revision and rewriting. Not enough control of element placement.* Participants report that they would look at their previous sentences to correct spelling mistakes and for inspiration to proceed with new sentences. Once they learned what descriptions were possible to create, participants report that they revised their sentences to improve their descriptions of the different elements on their planet (e.g. size and colour). Some participants report that the possible descriptions were too limited, meaning there was no need to revise and rewrite. They report attempts at rewriting, but felt they were limited by what the application offered. Participants also report a discontent with the lack of control over where elements were placed on the planet. The archive function was primarily used to look at previous sentences, and most participants would not rewrite directly in the archive.

7.3.5 Social

Themes are: *Neutral opinions. Would share planets if system offered more customization.* When asked whether they would want to share their planet with classmates, and whether others could find inspiration in their planets, participants reported mostly neutral responses. They report that they could be interested in sharing their planets with each other, but felt that the system is too limited in its current state to offer any purpose in sharing.

The themes derived from the qualitative replies show a general enjoyment of the *The Wor(l)d Builders* prototype application. However, the reports show that participants found the application to be too limited in its current state, especially concerning customization of their planets.

Chapter 8

Discussion

In this chapter we will discuss the results of the test, to uncover how to improve upon *The Wor(l)d Builders* application.

8.1 Test for Accuracy

In this section we will discuss the results of the Accuracy test.

The Precision of 73.1% meant that approximately 1/4 of the time an object was described that did not appear in the sentence correctly. Many of these errors were due to participants writing numbers rather than their numeral counterpart. This suggests that we might need to accept numbers as well, even though the intention was to have the participants practice spelling of the numeral words. The Recall of 84.7% was quite high, but there were some issues that will be described at length in Interpretation & Modeling (False Positives). True Negative Rate of 26.5%, did not reflect what was intended (i.e. Gibberish sentences and spelling errors). Many of the false negatives were a reflection of the errors described in Dependency Relationships (False Negatives). The Accuracy of 67.4%, we deem unsatisfactory. There were, however, a lot of false positives and false negatives that affected all the measures in the accuracy of the model, many of which we are able to correct ourselves. This does in our opinion effectively show prospects of better results in the future. In this section we will discuss why that is, and how we might improve the accuracy of the application.

8.1.1 Interpretation & Modeling (False Positives)

We humans perceive the world quite differently, from individual to individual. Therefore, one must be careful with leaving things open to interpretation, when creating an application intended as a learning tool. In *The Wor(l)d Builders*, there were two prominent examples. When some pupils

were describing the need for water, on their planet (which were in some cases interpreted as being an Island), they would use words like '*sø*' or '*flod*' (i.e. '*lake*' or '*river*'), none of which were used in the text prompts that were intended for *Modeling* (see Writing & Didactics section). The cause of this may lie in the visual representation of water, on the pre-built planets, which could be interpreted as a lake or river. If the visual representation influenced the pupils writing, it is quite interesting to know that the pupil did not only use the text for modeling, but also the visual representation of the texts. Some people may remember images better than texts, which means that regarding *The Wor(l)d Builders*, we should make sure that the visual representation adheres to the description in the text prompt, and we could link multiple nouns to the same object, using these as visual modifiers for the object in question (e.g. a '*river*' would modify '*water*' to be a long plane, and *lake*' as a spherical plane). There are of course many cases, not known to us yet, where such trivial solutions might not be valid. There were in fact some cases of the word '*sø*' that such a solution would not be able to handle. While the VISL parser is sufficient in classifying dependencies, word lemmas and part of speech, some issues with word ambiguity, within the context of *The Wor(l)d Builders*, arose. The word '*sø*' in its plural form is '*søer*', which is, in Danish, the same as the feminine form of the word '*pigs*' (i.e. '*sows*'). This meant that the VISL parser, in many cases, would interpret the word as '*sows*' rather than '*lakes*' and return the lemma '*sow*' ('*so*') instead of '*lake*' ('*sø*'). If a Dependency Parser was trained on data that suited the content of *The Wor(l)d Builders*, such word ambiguities may not arise, since the possibility of pigs, is not present in the application which would increase the accuracy of the model as less false positives would occur. We would however need to collaborate with Linguists to fully realize such a dependency parser. Furthermore, seeing as we aim to expand the design and effectively the application, we would need a method for dealing with such ambiguities because more possible objects (i.e. nouns), would mean more frequent word ambiguities.

There were also cases were the test participants wanted to spawn gameobjects there were not present in the application in any shape or form. For instance, one participant wrote *Der skal være en flot regnbue* (i.e. *There should be a beautiful rainbow*). It might be a cause of a participant wanting to test the systems limits, or a reflection of the goal being not stated clearly enough. Seeing as there were quite a few cases of such False Positives, we believe it is the first cause, but the design of the presented goal should be iterated on, and tested in isolation to make sure that it is perceived as intended.

8.1.2 Dependency Relationships (False Negatives)

The accuracy is also a reflection of an error in the data analysis of the output data from VISL. The analysis did not accept nouns of common gender in noun phrases, but only neuter gender. This error did however only occur when a noun had the dependency relation NPHR (Noun Phrase), and was of neuter gender. This is of course a reflection of our lack of expertize within the field of Linguistics, which is why we call for the collaboration with Linguists if one is to design for Natural Language Processing use, such as we have intended in, in an application for Game-based learning. That being said, the dependency relationships and genders we did include was found every time they occurred. In every case of them being a valid object, they were spawned on the participants' planet, highlighting the very prominent prospects of design, implemented in *The Wor(l)d Builders*. **Furthermore, it is evident that we need to decrease the discrepancy between text and visual object, in the modeling technique applied in the implementation.**

8.2 Questionnaire

As can be seen in fig. 7.1, the total means from the questionnaire are above average in weight, where 1 would be the lowest and 7 the highest. Participants tended to score the likert items above average, indicating semi-positive agreement with the statements. The lowest mean was 4.088 and the highest 5.98. When looking at the qualitative responses (i.e. the elaborations of the answers), we argue that participants generally enjoyed, or were at least intrigued by the concept, but had some reservations towards the limitations of the prototype application. An example of this can be found in a response from participant eight who said *'Not everything worked, but I feel that it (the application) was pretty good at figuring out what I wanted it to.'* as a reply to item 01. Participant ten, when answering the same question, reported: *'The things I wrote appeared on the planet. Though not all at once, and there were some things that I couldn't find at all'*. Generally, it seems that participants were happy when their trees and rocks worked as intended, but were discouraged by the limited amount of available objects, **and sentence structure**. As mentioned in the Dependency Relationships (False Negatives) section, many *false negatives* occurred when participants wrote about '*Rocks*'. This may be one of the causes of this discouragement, as rocks were clearly one of the objects that participants could explore and interact with on the explorable planets. Participants were told that their task was find elements of the environment on the other planets. Elements that they would like to be on their own planet. Nevertheless, some participants attempted to spawn objects that did not appear on the other planets, such as rainbows and volcanoes. Participant 15 rated item 01 with a 5 out of 7, and stated:

'It worked pretty well (the application), but you have to be very precise in what you are writing (sentence structure)'. The statement can be related to both limitations in sentence structure and spawnable objects. However, such limitations does not seem to discourage the participants to an extent were he thinks the application did not make his text come alive. We emphasized the utilization of the other planets for inspiration. Therefore, it seems reasonable that participants felt confused in cases where objects that appeared on the other planets did not seem to be able to spawn on their own planet. The utilization of other planets for inspiration, was, nevertheless, the highest rated items. Qualitative responses support this notion, with a general consensus of the planets being used multiple times for inspiration and for spelling correction (e.g. *'that was how I found inspiration for my own planet, and to formulate my descriptions'* - participant 3, *'I returned multiple times to the other planets, to find inspiration for e.g. specific trees* - participant 31). One could argue that the utilization of the planets, and the high rating of the items, stem from the participants being explicitly told to explore these planets. Modeling within *The Wor(l)d Builders* application occur on the other planets, and the results indicate them as serving their purpose well. Some participants report that planets were too empty, and did not have enough content on them (*'there wasn't enough on the planets for one to go back to them'* - participant 6, *'it wasn't necessary (to go back), since there weren't that many different things'* - participant 9). When observing the experiment, we noticed that many participants attempted to interact with all of the objects found on the snow planet, while only the snow covered trees were actually interactable. Descriptions of the objects on the different planets, all concerned information about the object, and its part in an ecosystem. We observed many participants attempting to write the exact same descriptions, or similar, for their own planet. An issue arose when these descriptions were phrased in ways that the system did not interpret properly. The modeling did not work as intended, because the text prompts were meant as inspiration that could not be copied directly. This was of course an error that needs to be corrected, as the participants did not use them as intended. Even with these issues, using the other planets for inspiration was still the highest rated item. One could argue that this was simply due to the planets being the only present modeling technique, even if this did not work entirely as intended.

The lowest mean was found in the items concerning whether participants wanted to share their text and planet with classmates. Participants responded with a general interest in the prospect of this form of social interaction in the application, but thought that the application was limited in its customization, thus removing the purpose of sharing the planet (*'If you could add more things, maybe even change the shape of the planet'* - participant 8, *'It could be fun to see how you have created your planet, compared to your classmates'* - participant 23, *'With a better formulated version (of my*

planet) and a more refined version of the game, it could be quite possible, but as of now, it (my planet) doesn't inspire great thoughts - participant 3). Their qualitative replies provide a much better understanding of their feelings towards the subject, explaining the total mean of 4.088. Participants are neutral towards the items, because, as can be derived from their explanations, they did not think the content was sufficient at the current state of the application but saw promise in its future development.

Participants agreement with items related to self-evaluation was fairly neutral (item 02 and item 03). As mentioned in the Results chapter, participants revised and rewrote their texts to various degrees. Those who did, did so usually to correct spelling mistakes and to find some inspiration in previous sentences (*'I cannot spell, so I had to return and look at my spelling mistakes'* - participant 7, *'I had to see what I had written before, to get inspiration'* - participant 26). It seems that some participants found revision and rewriting to be unnecessary due to the sentence construction and limited spawnable objects, and would opt to go back and visit other planets, or simply attempt a new sentence instead (*'I didn't bother, and instead attempting to write other things'* - participant 15, *'You learn quickly which descriptions are easily accepted'* - participant 28). Other participants used the function to evaluate on their previous sentences and improve their writing and/or change their planet (*'I used my previous input to improve my method of writing my sentences'* - participant 27, *'I changed my sentences multiple times to improve my planet'* - participant 31). Some participants complain about the lack of control of the planet, especially concerning object placement (*'The issue was that I couldn't decide where on my planet my thing would spawn'* - participant 15, *'You should have the possibility to place the trees yourself, and so on'* - participant 12). Finally, some participant reported never revising and rewriting, forgetting the archive function, or only doing it once (*'I didn't rewrite'* - participant 18, *'I kind of forgot to look in the archive'* - participant 9, *'I checked once to rewrite it, so that it understood that I wanted water and not a system, but other than that, no'* - participant 5). As can be seen from these examples, the qualitative responses reflect the neutral agreement value. Participants had mixed feelings towards self-evaluation, and the feature was utilized to a varying degree. These responses are very interesting when thinking of the future of the application. The formative evaluation of the application should encourage these self-evaluations, which should eventually lead to improved writing capabilities. Participant 27 reflected this in his reply, where the participant claims to use their previous input to improve their sentences. This shows a proper reflection and revision of their own work, and thus shows potential for the kind of self-evaluation, that we hope to achieve with the application. An increase in customization and control, more spawnable objects, and more focus on the archive as a functionality seems to be what could greatly encourage more self-evaluation.

Item 06, 07, 08 and 09, did not include any qualitative responses. The means can be interpreted as participants somewhat agreeing to the possibility of the application being used for education in middle school. The means being above neutral, but not reaching full agreement, can be a reflection of participants general feeling towards the application. They might not see it as a feasible educational tool in its current state, but can imagine a refined version of the application being used.

8.3 Questionnaire Comparison

In the Results chapter we compared the likert items between high school students and middle school children, and found a significant difference between item 06 and item 08. In the test with unequal participants, the significance proved to be a medium effect size, while in the test with equal participants, this significance was closer to a large effect size. When comparing all items between both groups, we found a significant difference with a small effect size. The results indicate that the difference in total items is primarily found in item 06 and 08, since these have approximately large effect sizes. Item 06 states *'The visual representation of my text, could motivate me to improve upon my writing'*. It is difficult to tell why the significant difference exists, but it could be due to a gap in understanding of the question itself. It might be that middle school children, simply did not understand the meaning of 'the visual representation', or maybe what it means to be motivated to improvement. It could be that they understood the question differently, or that they just did not agree at the same rate as high school students. The same could be said for item 08, which states *'I would receive a better understanding of a given subject, had I been taught the subject through the game'*. It might be that high school students can better relate to the questions, due to their maturity and more refined metacognitive skills. High school students have already been through middle school, and the question offers them a chance to look back and reflect on whether they would have learned more if they had been taught through the game. We do not know whether the application in itself was too complicated for the middle school children, which then created the difference, or if they simply did not like the application as much. The mean ranks for the equal participants test support this, by showing every mean rank being higher or equal for high school students, even if these were not significantly different. Besides difficulties with spelling, we did not observe any increased difficulty in usability between the two groups, during the experiments. When looking at the logged data, what seemed to be the most prominent difference between the two groups, was that the participants from middle school had more difficulty with spelling. We argue that this could be discouraging, due to the extensive time spent on correcting spelling mistakes, instead of receiving visual output. In the Design chapter,

we mentioned the idea of a 'notebook' feature, where pupils could note the different objects that they had discovered. The notebook could serve as an extension of the modeling techniques, where middle school children can see how words are spelled once they get back to their own planet. The notebook feature, and a method of simplifying the application could be what would enhance middle school children's experience in the application.

Seeing as many of the errors were caused by an error in the data analysis of the VISL output. Errors that we know how to fix, we deem the VISL parser, and more specifically Dependency Parsing and lemmatizing, quite suitable for the purpose of *The Wor(l)d Builders*. The modeling techniques that were applied worked to some extent, as the participants '*learned*' what types of objects they could include in the creation of their own planet. There were, however, issues with the object descriptions not reflecting the actual sentences that pupils could write, which contradicts the purpose of modeling. Generally, participants seemed to enjoy the application, but were, sometimes, discouraged by the lack of customization, that the application had to offer. Furthermore, the application might have been too difficult for the middle school pupils, or perhaps they simply did not find it as enjoyable as the high school students. We are left with a somewhat satisfactory results, as most of the issues that we deem to have affected our results negatively, should be easily fixed. We have become aware of the rest of these issues, and will work towards finding solutions. The future of *The Wor(l)d Builders* is promising.

Chapter 9

Future Works

This chapter will present our thoughts and ideas on the improvement of the application, based on the discussion of our test results.

9.1 Application changes

There are a number of immediate changes that can be implemented based on the results.

Modeling 3D objects will be changed to properly match their text description. Furthermore, text descriptions will be rewritten to better suit the purpose of modeling. More 3D models will be added to the different planets, to increase customization options. A notebook feature will be added, which allow pupils to note their observations of the objects and descriptions on other planets. This notebook can then be used as a reference for spelling and personal modeling, when writing for your own planet.

Home Planet More adjective options will be implemented, which will increase the customization of the home planet. Furthermore, object placements will be altered to enable pupils to have more control over where their objects spawn on the home planet. We will find a method of implementing water as actual water, instead of symbolizing it with clouds.

UI Input field in home planet scene should be larger. Archive feature should be changed to work properly, and should be emphasized more as a feature.

9.2 Natural Language Processing changes

In order to expand upon the vocabulary of available nouns, adjectives and to include prepositions etc. we need to allow for more dependency relation-

ships between words. A prepositional dependency could for instance allow pupils to place objects at a given location on the planet, rather than being spawned randomly. Furthermore, we could include verbs to bring some actions to the application (e.g. raining clouds). This should encourage more elaborate sentence structures, and allow for more customizable planets.
If we are to train our own dependency parser, we would need to collaborate with Linguists to produce annotated linguistic data, for the purpose of training the parser.

9.3 The Wor(l)d Builders

This thesis focused on developing the functionality of the *The Wor(l)d Builders* application, both in regards to didactic design, and the natural language processing implementation. The future of the application is that of a proper educational game with a focus on the development of ludic elements, once the main functionality works as intended. In the future, it would interesting to conduct a long term experiment with middle school children, to find if *The Wor(l)d Builders* could exceed regular classroom methods of improving writing capabilities.

Chapter 10

Conclusion

In this project we designed *The Wor(l)d Builders*, an application meant to motivate Danish middle school pupils to write, and increase their writing capabilities. To that end we conducted research analysis in the fields of Didactics, Game-Based Learning and Natural Language Processing. The purpose of this Thesis was to find the best possible methods from the research fields, to include in the Design of *The Wor(l)d Builders*. A prominent natural language processing method, called Word2Vec, was trained and implemented. The implementation of the model was however not satisfactory, and it was deemed unsuitable for the purpose of this thesis. We found that collaboration with Linguists is vital in designing for educational games, when utilizing Natural Language Processing. Specifically, we need to collaborate with Linguist to create annotated data, to train a dependency parser. We got access to Visual Interactive Syntax Learning tool's dependency parser API, and implemented it in *The Wor(l)d Builders* application. The output of VISL segmented into commands that visualized the input sentences in the application. From the didactic research, we created a model called *The Writing Cycle*, on which we based the didactic design. An experiment was conducted in an attempt to measure the accuracy of the application and to get insight into the participants thoughts about the design, and its prospects. Results showed that the implementation successfully visualized objects being described 67.4% of the time. A percentage the describe the extent to which we were able to '*teach*' *The Wor(l)d Builders* application to interpret Danish texts and visualize said texts. Furthermore, qualitative data and observations showed that participants enjoyed the application, but had some reservations regarding it limitations at its current state. We are satisfied with the results, as we are aware of the various issues found during the experiment. We can now work towards expanding upon the design, and implementing more features, and fine tuned methods that we believe will encourage middle school children to improve their writing capabilities.

Chapter 11

Bibliography

[1] Agus, J. R. and Winiharti, M. (2011). The analysis of creative writing teaching through story book reading for the first grade students of tunas muda international school. *Lingua Cultura*, 5(2):98.

[2] Al-Rfou, R., Perozzi, B., and Skiena, S. (2013). Polyglot: Distributed word representations for multilingual nlp. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 183–192, Sofia, Bulgaria. Association for Computational Linguistics.

[3] Alberti, J. (2008). The game of reading and writing: How video games reframe our understanding of literacy. *Computers and Composition*, 25(3):258–269.

[4] Apostolellis, P. and Bowman, D. A. (2016). Audience involvement and agency in digital games: Effects on learning, game experience, and social presence. In *Proceedings of the The 15th International Conference on Interaction Design and Children*, IDC '16, pages 299–310, New York, NY, USA. ACM.

[5] Barab, S., Pettyjohn, P., Gresalfi, M., Volk, C., and Solomou, M. (2012). Game-based curriculum and transformational play: Designing to meaningfully positioning person, content, and context. *Computers & Education*, 58(1):518 – 533.

[6] Bates, A. W. (2015). *Teaching in a Digital Age*.

[7] Becker, L. (2012). Discuss: Toward a domain independent representation of dialogue.

[8] Benson, A. D., Joseph, R., and Moore, J. L. (2017). *Culture, Learning, and Technology: Research and Practice*. Taylor & Francis.

[9] Berzins, E. (2018). Ultra sci-fi game audio ambience pack vol. 1.

[10] Bird, S., Loper, E., and Klein, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media.

[11] Brand, J. E. and Todhunter, S. (2015). Digital australia 2016.

[12] Brok, L. S., Bjerregaard, M. B., and Korsgaard, K. (2015). *Skrivedidaktik: en vej til læring*. Literacy & læring. Klim.

[13] Chowdhury, G. G. (2003). Natural language processing. annual review of information science and technology, 37. pages 51–89.

[14] Colby, R. S. and Colby, R. (2008). A pedagogy of play: Integrating computer games into the writing classroom. *Computers and Composition*, 25(3):300–312.

[15] Csikszentmihalyi, M. and Csikszentmihalyi, I. (1992). *Optimal Experience: Psychological Studies of Flow in Consciousness*. Cambridge University Press.

[16] DARTH$_A RTISAN$(2017).$Freetrees$.

[17] Davis, K. D. (2017). *The Effects of Play-Based Learning On A Child's Social, Emotional, and Academic Learning*. PhD thesis. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2018-03-02.

[18] Doly, A.-M. (2007). Metacognition to learn how to write texts at school and to develop motivation to do it. In [38], pages 381–392.

[19] Duh, E. S., Koceska, N., and Koceski, S. (2016). Game-based learning: educational game azbuka to help young children learn writing cyrillic letters. *Multimedia Tools and Applications*, 76(12):14091–14105.

[20] Eagle, M. and Barnes, T. (2009). Experimental evaluation of an educational game for improved learning in introductory computing. *ACM SIGCSE Bulletin*, 41(1):321.

[21] FACTORY, L. R. S. (2015). Ui sfx.

[22] Ferdig, R. and Winn, B. (2009). The design, play, and experience framework. pages 1010–1024.

[23] Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-Learning and Digital Media*, 2(1):5–16.

[24] Kinzie, M. and R. D. Joseph, D. (2008). Gender differences in game activity preferences of middle school children: Implications for educational game design. 56:643–663.

[25] Kokkegård, H. (2014). Computerspil gœr bœrn bedre til engelsk.

[26] Kübler, S., McDonald, R., and Nivre, J. (2009). *Dependency Parsing*. Online access: IEEE (Institute of Electrical and Electronics Engineers) IEEE Morgan & Claypool Synthesis eBooks Library. Morgan & Claypool.

[27] Lika, M. (2017). Teaching writing. *IJAEDU- International E-Journal of Advances in Education*, page 152–152.

[28] Manning, C. D. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Mit Press. MIT Press.

[29] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.

[30] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. *CoRR*, abs/1310.4546.

[31] Norman, D. (1993). *Things that Make Us Smart: Defending Human Attributes in the Age of the Machine*. A William Patrick book. Addison-Wesley Publishing Company.

[32] Osunde, J., Windall, G., Bacon, L., and Mackinnon, L. (2015). An investigation of digital games features that appeal to young females and males.

[33] Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education*, 52(1):1–12.

[34] Paz, S. D. L. and Graham, S. (2002). Explicitly teaching strategies, skills, and knowledge: Writing instruction in middle school classrooms. *Journal of Educational Psychology*, 94(4):687–698.

[35] Rahimi, Z., Litman, D., Correnti, R., Wang, E., and Matsumura, L. C. (2017). Assessing students' use of evidence and organization in response-to-text writing: Using natural language processing for rubric-based automated scoring. *International Journal of Artificial Intelligence in Education*, 27(4):694 – 728.

[36] Rensing, J., Vierbuchen, M.-C., Hillenbrand, C., and Grünke, M. (2016). Implementing peer-assisted writing support in german secondary schools. *Insights into Learning Disabilities*, 13(2):151 – 164.

[37] Riddle, T., Bhagavatula, S. S., Guo, W., Muresan, S., Cohen, G., Cook, J. E., Purdie-Vaughns, V., and International Educational Data Mining, S. (2015). Mining a written values affirmation intervention to identify the unique linguistic features of stigmatized groups.

[38] Rijlaarsdam, G., Bergh, H., and Couzijn, M., editors (2007). volume 14 of *Studies in Writing*. Springer Netherlands.

[39] Roscoe, R., Snow, E., Brandon, R., and McNamara, D. (2013). Educational game enjoyment, perceptions, and features in an intelligent writing tutor. pages 515–520.

[40] S. Paras, B. and Bizzocchi, J. (2005). Game, motivation, and effective learning: An integrated model for educational game design.

[41] Shapira, Anat & Hertz-Lazarowitz, R. (2007). Metacognition to learn how to write texts at school and to develop motivation to do it. In [38], pages 381–392.

[42] Shulgach, A. N. (2018). 3d modern menu 1.

[43] Smidt, J., Solheim, R., and Aasen, A. (2013). *På sporet af god skriveundervisning: en bog for lærere i alle fag*. Klim.

[44] Squire, K. and Jenkins, H. (2011). *Video Games and Learning: Teaching and Participatory Culture in the Digital Age*. TEC series. Teachers College Press.

[45] STUDIOS, S. (2017). Polygon - viking pack.

[46] Svanes, I. K. and Skagen, K. (2016). Connecting feedback, classroom research and didaktik perspectives. *Journal of Curriculum Studies*, 49(3):334–351.

[47] Tesnière, L. (1959). *Éléments de syntaxe structurale*. Éditions Klinksieck.

[48] Undervisningsinisteriet (2016). Fagformål for faget dansk.

[49] Undervisningsinisteriet (2017). It i undervisningen.

[50] Wilson, J., Roscoe, R., and Ahmed, Y. (2017). Automated formative writing assessment using a levels of language framework. *Assessing Writing*, 34:16 – 36.

[51] Wouters, P. and van Oostendorp, H. (2017). *Overview of Instructional Techniques to Facilitate Learning and Motivation of Serious Games*, pages 1–16. Springer International Publishing, Cham.

...