FAKULTÄT FÜR !NFORMATIK

Faculty of Informatics

Testing Perceptual Preference

Development and Evaluation of Computer Based Methods for Testing Perceptual Preference

DISSERTATION

submitted for the degree of

Doktor/in der technischen Wissenschaften

by

Mag.phil. Mag.rer.soc.oec. Gudrun Salamon
Registration Number 00009896

to the Faculty of Informatics
at the Vienna University of Technology

Advisor: Ao.Univ.Prof. Dr. Dieter Merkl

The dissertation has been reviewed by

Ao.Univ.Prof. Dipl.-Ing. Dr. Gerald Futschek
Vienna University of Technology, Austria

Prof. Dr. Bettina Berendt, M.Phil. Dipl.-Kff. M.Sc.
Katholieke Universiteit Leuven, Belgium

Vienna, 01.11.2017

MMag. Gudrun Salamon

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Testmethoden zur Identifikation von Wahrnehmungspräferenz

Entwicklung und Evaluation computergestützter Methoden zur Testung von Wahrnehmungspräferenz

DISSERTATION

zur Erlangung des akademischen Grades

Doktor/in der technischen Wissenschaften

eingereicht von

Mag.phil. Mag.rer.soc.oec. Gudrun Salamon
Matrikelnummer 00009896

an der
Fakultät für Informatik der Technischen Universität Wien
Betreuung: Ao.Univ.Prof. Dr. Dieter Merkl

Diese Dissertation wurde begutachtet von

Ao.Univ.Prof. Dipl.-Ing. Dr. Gerald Futschek *Vienna University of Technology, Austria*	Prof. Dr. Bettina Berendt, M.Phil. Dipl.-Kff. M.Sc. *Katholieke Universiteit Leuven, Belgium*

Wien, 01.11.2017

MMag. Gudrun Salamon

Erklärung zur Verfassung der Arbeit

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

(Ort, Datum)	*MMag. Gudrun Salamon* (Unterschrift Verfasserin)

perception

per·cep·tion
/pəˈsɛpʃ(ə)n/
[per-sep-shuh]
n.

The process by which people translate sensory impressions into a coherent and unified view of the world around them.

For those who have influenced my view of the world:
you have made a difference in my life.

Abstract

One of the great challenges in user modelling is the question how to gain personal information about the user. Of course, the most direct way is to ask him or her. Nevertheless, this option is not much appreciated by the user: It takes time and effort to fill in questionnaires and the user might quickly get a feeling of being kept from his or her original goal. On the other hand, the user is not always aware of the information needed by a system, especially when it gets to complex aspects of his or her personality or preferences.

Perceptual preference can be considered as an aspect of cognitive style and signifies an individual's preference or preference pattern regarding perceptual channels of information presentation and information processing. This thesis develops and evaluates three different methods for testing perceptual preference.

The first one is an explicit method, eliciting perceptual preference via the Perceptual Preference Questionnaire (PPQ). Perceptual preference is assessed in regard to information processing, knowledge gain, and learning. It is analysed via the scales visual, auditory, kinaesthetic and olfactory/gustatory preference and examined in regard to distribution as well as co-occurrence patterns of perceptual preference and user interests.

The second method is an implicit method, investigating user-generated text as an indicator of perceptual preference. In order to study the use of sensory vocabulary within German text, we develop the Signal Term Extraction Method (STEM) Algorithm and integrate it into an analysis pipeline. The analysis of a huge forum corpus is focused on the use of sensory vocabulary patterns as well as response behaviour.

The third method is an embedded method. The Game Embedded Testing of Learning Strategies (GETOLS) method embeds the testing procedure into a didactic adventure game. The results obtained automatically by our test environment are compared to the ones elicited via conventional testing by a psychologist.

Overall, the evaluation results show a high potential for considering perceptual preference as a valuable extension to existing user models.

Kurzfassung

Eine der Herausforderungen im Bereich User Modeling liegt in der Gewinnung von persönlichen Informationen über User. Die direkte Befragung wird von den Usern nicht gerade geschätzt: Fragebögen auszufüllen bedeutet Aufwand und braucht Zeit, schnell entsteht bei den Usern das Gefühl, vom eigentlichen Vorhaben abgehalten zu werden. Darüber hinaus sind sich User nicht immer der Informationen bewusst, die von einem System gebraucht werden, besonders, wenn es sich um komplexe Aspekte von Persönlichkeit oder Präferenzen handelt.

Wahrnehmungspräferenz als ein Aspekt von Cognitive Style bezeichnet die individuelle Präferenz von Sinnessystemen, die bei der Präsentation und Verarbeitung von Information zum Einsatz kommen. Die vorliegende Arbeit entwickelt und evaluiert drei unterschiedliche Testmethoden zur Identifikation von Wahrnehmungspräferenz.

Die erste ist eine explizite Methode, Wahrnehmungspräferenz wird hierfür mit dem eigens entwickelten Perceptual Preference Questionnaire (PPQ), mittels Fragen zu Informationsverarbeitung, Wissensgewinn und Lernen erhoben. Es ergibt sich eine Unterteilung in die Skalen visuelle, auditive, kinästhetische und olfaktorisch-gustatorische Präferenz, die Ergebnisse werden im Hinblick auf Distribution sowie Kookkurrenz von Wahrnehmungspräferenz und Interessen der User untersucht.

Die zweite ist eine implizite Methode, welche User-generierten Text als Indikator für Wahrnehmungspräferenz erforscht. Zur Analyse des Wahrnehmungswortschatzes in deutschsprachigem Text entwickeln wir den Signal Term Extraction Method (STEM) Algorithmus und integrieren diesen in eine Analyse-Pipeline. Die Untersuchung eines großen Forenkorpus erfolgt mit einem Schwerpunkt auf Stabilität und Verteilung von sinnesspezifischem Vokabular sowie Antwortverhalten der User.

Die dritte ist eine integrierte Herangehensweise. Die Game Embedded Testing of Learning Strategies (GETOLS) Methode bettet die Testprozedur in ein didaktisches Adventurespiel ein. Die so erzielten Ergebnisse werden mit den in einem konventionellen Testsetting mit einem Psychologen als Ausführenden verglichen.

Die Evaluationsergebnisse sprechen für die Erweiterung von User Models um eine Komponente zu Wahrnehmungspräferenz.

Contents

1 Introduction **13**

1.1 Methodology 14
1.2 Contributions 16
1.3 Thesis Outline 18

2 Perceptual Preference **21**

2.1 Related Research 21
2.1.1 Preference and Perception 21
2.1.2 Cognitive Style 27
2.2 Relevant Contexts: User Modeling, Learning, e-Commerce 29
2.2.1 User Modeling 29
2.2.2 Learning 30
2.2.3 E-Commerce 32
2.3 Existing Test Methods 32

3 PPQ: Perceptual Preference Questionnaire **35**

3.1 PPQ: Test Design 36
3.2 Methodology 39
3.2.1 The Online Survey 39
3.2.2 Procedure 40
3.2.3 Data Description 40
3.3 Test Validation 41
3.4 PPQ and Interests 44
3.4.1 Distribution and Co-occurrence Patterns 45
3.4.2 Correlations 46

4 The Vocabulary of Perception **49**

4.1 Introduction 49
4.2 Related Research 50
4.2.1 Psycholinguistics 50
4.2.2 Using Text as a Source of Implicit Information on its Author 51
4.3 Vocabulary of Perception 53
4.3.1 Our Vocabulary of Perception 54
4.4 Lexical Analysis Algorithms 55
4.4.1 Background and Related Research 58
4.4.2 General Structure of the Algorithms 60
4.4.3 The Algorithms 62
4.4.4 Implementation 65
4.4.5 Evaluation 67
4.5 The Signal Term Extraction Method (STEM) Algorithm 72
4.5.1 Operations and Components 74
4.5.2 Execution Examples 78

4.6 The VAKOG Analysis Pipeline 80
4.6.1 Measures 80
4.6.2 Hypotheses 81
4.6.3 Data Preparation and Analysis Procedure(s) 84
4.7 The Vocabulary of Perception in Forums 87
4.7.1 E30 Forum Corpus 87
4.7.2 Results 88

5 GETOLS: Game Embedded Testing of Learning Strategies 91

5.1 Introduction 92
5.2 Related Research 93
5.2.1 How to Test Learning Strategies 93
5.2.2 Embedded Testing 94
5.3 EAGC: An Excursus to Educational Game Design 95
5.3.1 Conceptual Design 97
5.3.2 Game Design 98
5.3.3 Implementation 99
5.3.4 Testing 100
5.3.5 Validation 100
5.3.6 Project Management 101
5.4 Method 104
5.4.1 Testing Method 104
5.4.2 Measures 107
5.4.3 Hypotheses 107
5.5 The GETOLS Test Setting 108
5.5.1 The Didactic Adventure Game "Save the city" 108
5.5.2 Integrating the Testing Method into the Game 109
5.5.3 Implementation and Testing 111
5.6 Experimental Results 112

6 Discussion and Conclusion 117

6.1 Answers to our Research Questions 117
6.2 Limitations 120
6.3 Cross Connections and Outlook 121
6.4 Summary 123

References 127

Appendix 143

List of Figures 143
List of Tables 144
Acknowledgements 147

CHAPTER 1

Introduction

Nothing is in the intellect that was not first in the senses.
(Thomas Aquinas)

Digital media permits to present information in manifold ways. An adequate mode of information presentation helps the user with information processing, e.g. user A prefers to read new information whereas user B definitely prefers listening to it. Research shows that information can be more easily understood if its presentation is adapted to the cognitive preference of the user that needs to be reached. One aspect of cognitive preference consists in the preferred mode of perception. Such information might be of interest for every user model used in a setting where the user's interest needs to be captured and/or the user's process of information perception and organization shall be supported. Therefore, it might be an interesting extension for adaptive hypermedia, especially for the field of flexible content and interface design. Such personalization could be useful for a wide range of applications, including, but not limited to, e-commerce and e-learning.

But how can such a preference be elicited? In the field of user modeling, there exists a huge number of models and systems that perform very well with only few or even without any direct interaction with the user in form of extensive questionnaires or tests. It would thus not make much sense to suggest a new variable for user models that needs intense interaction with the user for being elicited. For this reason, this thesis develops an explicit, an implicit and an embedded testing method for testing perceptual preference. Therefore, we first present the Perceptual Preference

Questionnaire (PPQ), which is designed as a twenty items questionnaire. Second, we propose to use existing text, written by the target user and published in the internet, as a resource of information on his or her perceptional preference, and to use natural language processing techniques for text mining and lexical analysis. Third, we develop the Game Embedded Testing of Learning Strategies (GETOLS) method, which embeds a learning strategies test within a didactic adventure game.

1.1 Methodology

Three different test environments are set up in order to investigate different test approaches: while the first one uses explicit testing, the second one focuses on implicit and the third one on embedded testing.

From the methodological point of view, this thesis follows the research paradigm of design science [Hevner et al. 2004; Gregor and Hevner 2013]. As an empirical research approach, we use explicit and implicit data gained by some qualitative, but mainly quantitative research. Computer science, as a basis and overall perspective, was enriched by knowledge and methods coming from educational science, psychology, and linguistics. By combining several testing approaches, we follow the tradition of mixed methods research, which is especially fruitful for interdisciplinary research [Johnson and Onwuegbuzie 2004].

The design science approach provides a framework for the development of new ideas by means of creating prototypical artifacts and the evaluation of those artifacts. Such artifacts should “extend the boundaries of human problem solving and organizational capabilities by providing intellectual as well as computational tools” [Hevner et al. 2004]. The framework situates information systems research within a problem space, defining not only the development and evaluation of the artifacts themselves, but also the research environment and the knowledge base that covers theories and methods with its foundations and methodologies. Research and design of artifacts should thus be motivated by relevance and be grounded on rigor [Wieringa 2014; Vaishnavi and Kuechler 2015]. An overview figure of that framework is illustrated in Figure 1.

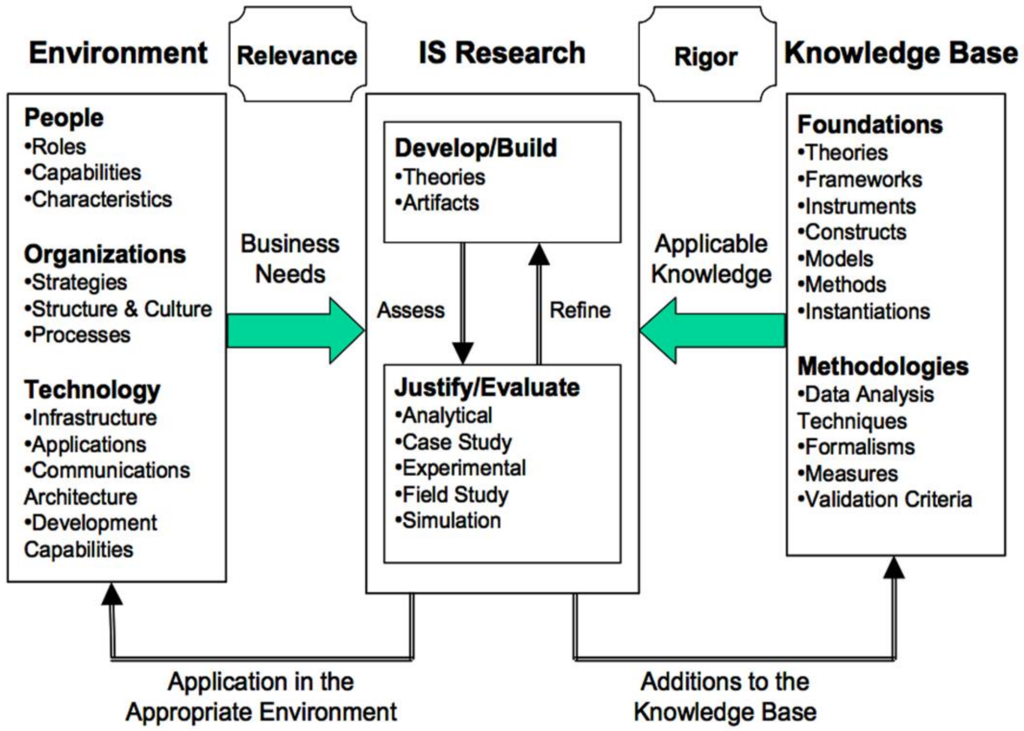

FIGURE 1 THE INFORMATION SYSTEMS RESEARCH FRAMEWORK [HEVNER ET AL. 2004]

From this point of view, the test methods developed in this thesis are artifacts that can be considered as *methods* whereas their implementations are *instantiations*. Furthermore, this thesis sums up existing knowledge and methods that are used as a basis for the development of our own theories and artifacts and also takes possible application domains into consideration.

For the evaluation of these artifacts we chose different validity approaches. The explicit testing method is validated via statistical scale analysis, the implicit testing method uses statistical analyses for hypothesis testing, and the embedded testing method uses repeated measures design. An overview of our research methodology can be found in Figure 2, its components are described in detail in the Chapters 3-5.

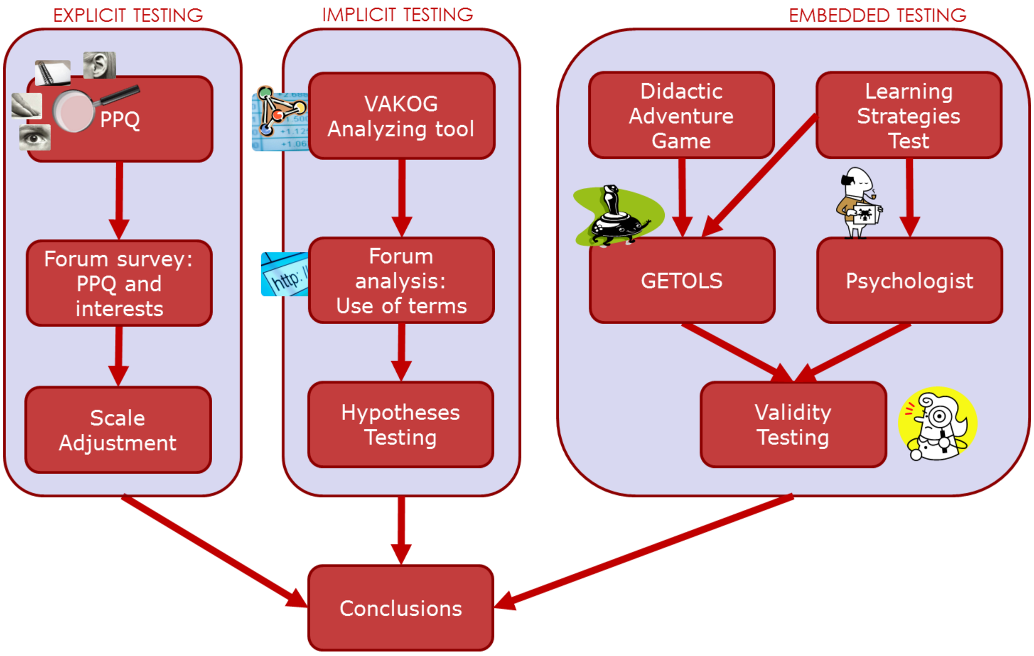

FIGURE 2 OUR RESEARCH METHODOLOGY

1.2 Contributions

The information systems research framework offers a model to classify contributions with the help of the knowledge contribution framework [Hevner 2011]. The two factors application domain maturity and solution maturity lead to four types of research, namely routine design, inspiration, exaptation, and invention. It is a known issue that the research field of user modeling seeks to develop new methods for embedded and implicit testing of user preference. Our methods to test such preference can hence be classified as inspirational research (by providing new solutions for known problems). The focus of this thesis, the combination of perceptual preference and computer science in order to test perceptual preference, is new and can hence be classified as a contribution by invention (by providing new solutions for new problems). The classification of our contributions according to the knowledge contribution framework is depicted in Figure 3.

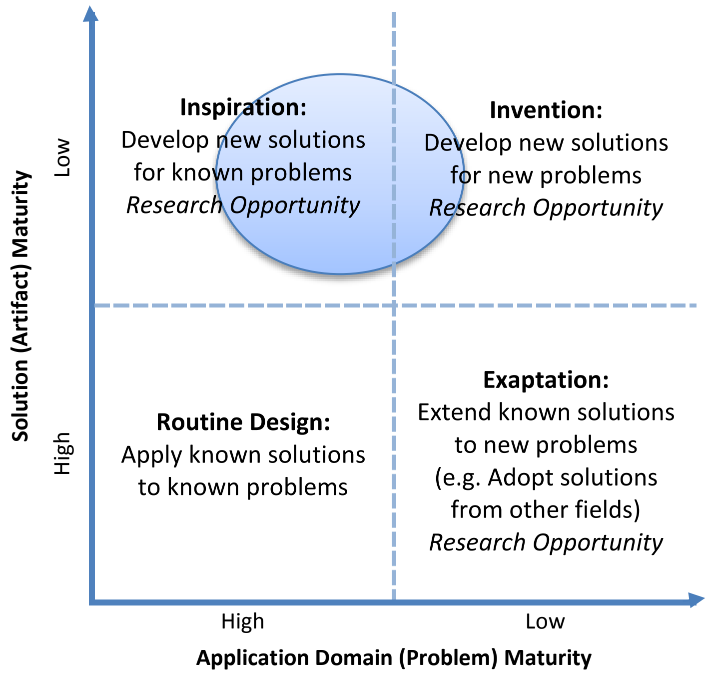

FIGURE 3 CONTRIBUTION CLASSIFICATION ACCORDING TO THE KNOWLEDGE CONTRIBUTION FRAMEWORK [HEVNER 2011]

In detail, the contributions of this thesis are:

Contribution 1: The Perceptual Preference Questionnaire (PPQ). Even though mainly developed as a verification method, the PPQ can be used as a valid elicitation tool for direct testing of perceptual preference via questionnaire. It consists of 20 test items, which load on the scales visual, auditory, kinesthetic and olfactory/gustatory preference. On a 5-point-Likert scale, participants can indicate their agreement with each test item on answer options ranging from “strongly disagree” to “strongly agree”.

Contribution 2: The VAKOG Analysis Pipeline. This method permits to extract sensory vocabulary from any text and to automatically build perceptual profiles upon the lexical analysis of expression of a user. Not only the pipeline itself, but also several important constituents that have been developed in the course of this thesis can be considered as a contribution. Besides the extension of the lexicon of sensory vocabulary and a refinement of the word grammar description model, the development of the Signal Term Extraction Method (STEM) algorithm counts to the core contributions within the VAKOG Analysis Pipeline.

Contribution 3: The EAGC Guidelines. The Guidelines for Educational Adventure Game Creation (EAGC) are a practical recommendation on how to realize the concept of serious games within the genre of adventure games. The guidelines address all relevant aspects of design, implementation and testing. In five phases, it is indicated which aspects to consider, what decisions should be taken and what dependencies to think of.

Contribution 4: The GETOLS Method. The Game Embedded Testing of Learning Strategies (GETOLS) method applies the method of embedded testing and adapts the test method of Vester's learning strategies [Vester 2011] in order to embed them into a didactic adventure game. The outcome of playing the game is, in addition to the didactic side effect, a test result on the personal learning strategies preference in the dimensions of visual, auditory, kinesthetic and text based information processing.

1.3 Thesis Outline

The thesis is structured as follows:

Chapter 2 explores the concepts of perceptual preference and cognitive style, presenting related research and existing test methods within the field of perceptual preference. User modeling, learning and electronic commerce are investigated as possible application domains.

Every chapter from 3 to 5 describes one method development and evaluation process and raises three research questions each, which are answered in chapter 6.

Chapter 3 develops the Perceptual Preference Questionnaire (PPQ). An analysis of existing test methods shows the necessity of a questionnaire focused on perceptual preference only and explains the composition of test items. A study with 76 participants is used for test validation. A second part investigates the relation between PPQ results and the user's interests. It searches for result overlaps in order to identify aspects where explicit testing might be replaced with knowledge about the users' interests.

Chapter 4 focuses on the vocabulary of perception and develops an analysis pipeline to extract information about perceptual preference from a text. We present the

Signal Term Extraction Method (STEM) Algorithm for German language, which was developed to identify the vocabulary of perception within text. The algorithm is embedded into the VAKOG Analysis Pipeline, which allows analyzing forum posts with regard to the use of sensory vocabulary. The VAKOG Analysis Pipeline is applied to user generated text taken from online discussion forums. We examine the use of sensory vocabulary in regard to its overall distribution, its frequency, the consistency of vocabulary written by a user and the stability of patterns.

Chapter 5 develops GETOLS, a method for Game Embedded Testing of Learning Strategies. As a basis, we present the Guidelines for Educational Adventure Game Creation (EAGC). We identify the most suitable learning strategy test for application in a game and find solutions how to adapt the test setting to the possibilities of an adventure game. The results elicited via GETOLS method are compared to the results from a conventional test conducted by a psychologist.

Chapter 6 discusses the methods developed in this thesis as well as their validation results. It considers them in the context of related research and provides a critical review as well as a summary of the thesis work. It recapitulates design and validation steps and results and sums up the thesis' contributions. As a final point, we explore opportunities for future work.

Most of the work of this thesis was presented at scientific conferences and published in conference proceedings. The headlines of chapters or sections related to such content are marked with a footnote referring to the respective publications. For nearly all papers, Gudrun Kellner (the author, her name changed to Gudrun Salamon after marriage) is the first author. The only exception is the result of a master thesis co-supervised by the author.

CHAPTER 2

Perceptual Preference

Each mind perceives a different beauty.
(David Hume)

This chapter provides an introduction to perceptual preference. It gives an overview of related research and explores existing test methods for perceptual preference. We also look into relevant contexts, which are user modeling, learning and electronic commerce. In this chapter, we only discuss related research and general approaches to perceptual preference that can be considered as a basis for the whole thesis. Furthermore, each chapter dealing with a new topic comes along with a section on related work relevant to the specific topic and its methods and challenges.

2.1 Related Research

This section explores related research. First, we provide an introduction into preference and perception and how those two concepts are combined. Second, we give an overview of cognitive styles, especially those that reflect upon sensory preference.

2.1.1 Preference and Perception

Following the definition of the Oxford Dictionary, preference can be interpreted as "a greater liking for one alternative over another or others" [Oxford Dictionaries 2010].

Another definition details the first part and describes preference as "the act, fact, or principle of giving advantages to some over others" [Merriam-Webster 2003]. The second definition seems to be more helpful in our context: While "a greater liking" refers to a highly subjective attitude that might be influenced by a number of irrational factors, "the act, fact, or principle" differentiates between several cases and decision drivers. Whereas "the act" itself gives information on the user's status who is about to decide something, "the fact" describes a latter status where the decision is already taken. The probably most interesting aspect for this thesis lies in the concept of a "principle of giving advantages to some over others". It refers to the idea that some preferences are not only unstable and timely decisions dependent on moment and circumstances they are taken in, but that there exist some long-term factors like beliefs and personality traits, which influence decisions and which can be referred to as more or less stable preferences.

Preference can not only concern tangible objects or items out of a list of options, but also intangible objects like ideas or beliefs and also more or less conscious neurologic or cognitive processes. Furthermore, by far not every preference, and even less all factors deciding a certain preference are conscious [Kahneman and Tversky 1982].

Such preference(s) need to be modeled in order to allow software to determine how interesting an item or an option is to a specific user. The preference model of a user can hence "be used to select and prioritize items that may be interesting to that user" [Dastani et al. 2001].

This thesis focuses on preferences in the field of perception. Perception can, in short, be defined as "the ability to see, hear, or become aware of something through the senses" [Oxford Dictionaries 2010], or, with more detail, as "(1) *a*: a result of perceiving (observation), *b*: a mental image (concept); (2) *a***:** awareness of the elements of environment through physical sensation, b: physical sensation interpreted in the light of experience; (3) *a***:** quick, acute, and intuitive cognition, *b***:** a capacity for comprehension" [Merriam-Webster 2003]. The second definition gives an overview of the various aspects that need to be taken into account when talking about perception. They can be split up into three stages of perception processing: (2) describes the input phase where physical sensation (and perhaps some bias based on the individual's experience) leads to a certain information flow towards an individual's cognitive system. Such information is then stored as described in (1) as a memory, either as a single observation or as a part of a mental image. Those two stages influence each other as well as (3) the individual's cognition and its "capacity for comprehension".

Perception is strongly linked to the sensory systems. Psychology differentiates between sensation and perception. While sensation helps to detect facets of a stimulus such as color, form or smell, perception deals with the recognition of the stimulus itself as a specific object by movement, location, background etc. From this point of view, detecting the color red is a sensation, detecting a strawberry is a perception. [Martin 2008]

The human sensory system consists of the human sense organs, which lead information to different brain areas that communicate with each other in order to complete the internal picture of reality. Our sensory systems are receptors of the environment, based on them we build our understanding for a tangible and comprehensible world [Frings and Müller 2013]. The senses can be considered as an individual's link to the outside world and are input channels for situations and actions from the environment. When storing something in the memory, and hence during any learning process, it is necessary to create inner representations of outer facts, situations, relations and actions. Even direct sensory input needs to be processed by the cognitive system first in order to leave an impression in the memory. Therefore, the act of sensation and perception can be considered as a mental transcription of reality into a format processable and storable by the human brain. From this viewpoint, colors, sounds, neural sensations, odors and tastes are mental constructs that are built in our mind upon sensory signal processing. The outer world can thus only be understood by individual reconstruction [Forcht 2009]. This interpretation is one of the basic postulates in constructivism and can be summed up in the idea that our environment, as we perceive it, is an individual or supra-individual construct [Foerster 1985].

Mental representations are the basis of knowledge storage and are influenced by individual states and traits [Schmidthals 2005]. Representations in our mind can be seen as mental images and reflect, analogously to models or maps, reality. However, every model is based on reduction of complexity and only captures a part of the picture. Such models are created individually or supra-individually, there consequently exists a huge variety of different models. Perception becomes hence something highly individual. A subjective perspective of experience is built upon an individual's experience(s) perceived by sensory input, which is determined by phenomenological and qualitative characteristics of the perceived situations and actions. Mental representations "enable a creature to respond to features of the world that are not immediately present, to use past experiences to shape present behaviour, to plan for the future, and, in creatures such as ourselves, to be sensitive to very abstract features of the world." [O'Brien and Opie 2004] A subjective estimation on the question whether a perceived content is based on an external stimulus or on an interiorly constructed idea is not necessarily corresponding to measurable facts [Heckmann and Esken 1998].

In cognitive science, the internal models of the outer world are referred to as representations. According to Heckmann and Esken [1998], all conscious subsystemic situations and actions are representations and stand for something: either, as in the case of sensory representations, for things, events, or situations and their phenomenological characteristics experienced via our senses, or, as in the case of propositional representations, for perceived circumstances. Representations mediate biochemical (physiological) materiality of a psychological event and the consequences of a physiological stimulation. Herrmann [1993] differentiates sensory representation for the perceived system environment, motoric representation for the perception and understanding of movement patterns and abstract-conceptual representation for linguistic and non-linguistic knowledge. Verbal and nonverbal representations are stored separately and linked with each other by object-expression relations [Paivio 2008]. The use of language allows us to make nonverbal representations accessible for and communicable to our environment.

Overall, perception can be considered as the most important input channel of any living creature. However, external information that needs to find its way into a cognitive system is filtered by perceptual limits because "perception is an act, and like all behavioral acts, it will have its limitations and will sometimes be in error" [Coren 2003]. Therefore, information gathered by our senses is filtered and hence inevitably incomplete [Frings and Müller 2013]. Such limits can be quite different from species to species (e.g., a dog hears sounds in a broader frequency range than humans, and insects have another concept of vision because of their facet eyes), but also vary within one species from one individual to another. Even among people without health or other restrictions, perceptual ability, perception processing and the perceived quality and intensity of a perceptual stimulus vary depending on the individual's physical limitations as well as the individual's awareness threshold in regard to, e.g., attentiveness or tiredness. Even though the sensory systems are continuously gathering information, only input exceeding a certain minimum level of intensity, the so-called stimulus threshold, will make it into consciousness [Kellermann 1997; Laming 1997]. Moreover, sensory input is filtered according to limitations of the human memory. One important factor for the act of perception is attention. During the act of perception, some parts of the perceived information are considered as more important or relevant as the rest of it, due to interests, preferences and/or focus. Attention can be considered as focalization and concentration of consciousness [Moray 2017]. Perception is thus not only a passive reaction to external stimuli, but may as well be an active and target-oriented act [Harm 2000]. In several steps, the processed information is reduced according to the attentiveness of the perceiver and the (individually defined) relevance of the perceived information [Anderson et al. 2013]. Those physical sensory limitations and information processing limitations influence the individual's view on the world as depicted in Figure 4.

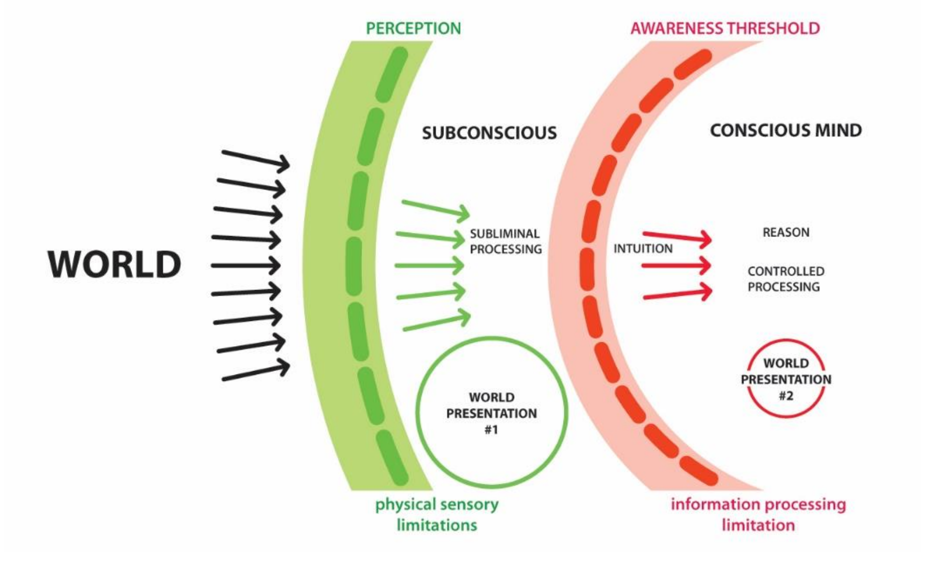

FIGURE 4 PERCEPTUAL LIMITATIONS DURING INFORMATION PROCESSING [SCHINDLING 2012]

Furthermore, the applied filters and hence the perceived content are also influenced by an individual's previous experience, beliefs, and expectations. Therefore, also an individual’s decisions and actions are influenced by his or her beliefs and expectations and their corresponding real-world experience, which is experienced based on perceptual stimulus. The overall result of perception is represented in the individual probability structure of the environment. This perception-action loop is depicted in Figure 5.

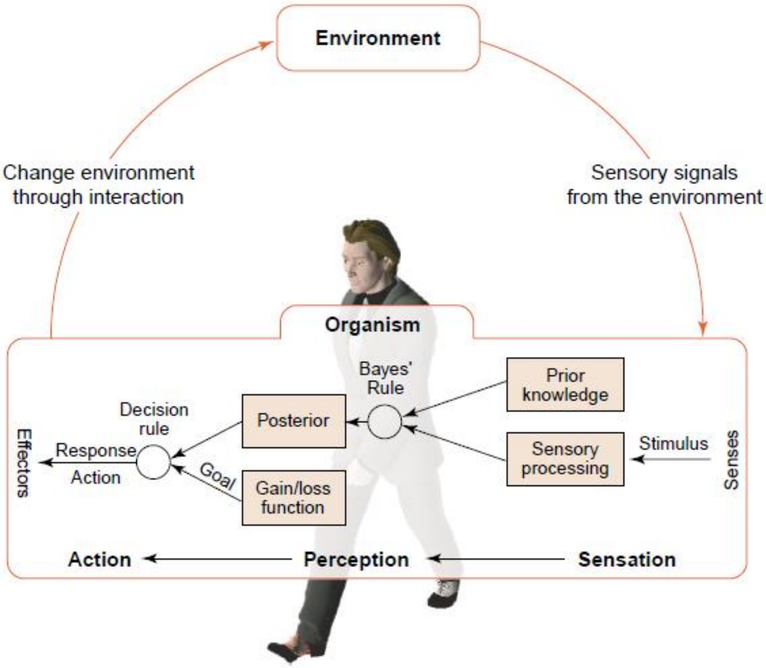

FIGURE 5 THE PERCEPTION-ACTION LOOP [ERNST AND BÜLTHOFF 2004]

„Our sensory systems are some of the most complex in the body and each allows us to experience the world in its rich, frequently drab, occasionally challenging form.“ [Martin 2008] The measurement of sensation poses a challenge to psychology and medicine: As it is highly subjective, it is hard to define scales in order to compare sensational impressions. Furthermore, there is no real agreement on the number of perceptual respective sensory systems one should take into account in the field of perception research [Gregory and Colman 1995; Laming 1997; Coren 2003; Goldstein and Brockmole 2016]. Wolfe et al. [2006] structure their work around the well-established model of the five senses, namely vision, audition, touch, olfaction and gustation. Goldstein and Brockmole [2016] describes vision in the aspects of perceiving objects and scences, motion, color, depth and size, audition in the aspects of localizing sound, perceiving speech, pitch and auditory scene, furthermore the cutaneous senses as well as the chemical senses. The combined processing of input via several sensory systems is described by Calvert, Spence and Stein [2004]. We decided to use the most common model of “the five senses”, which includes the systems of vision (seeing), audition (hearing), kinesthetics (feeling in means of tactile input as well as body sensation and emotions), olfaction (smelling), and gustation (tasting). We can hence define: sensory systems S = {visual, auditory, kinesthetic, olfactory, gustatory}.

Bringing both of these terms together, perceptual preference[1] means an individual’s preference or preference pattern regarding perceptual channels of information presentation and information processing, in our case visual, auditory, kinesthetic, olfactory, and gustatory.

2.1.2 Cognitive Style

Cognitive style “may be defined as an individual’s consistent approach to organising and processing information during thinking” [Sadler-Smith and Riding 1999]. In education, cognitive or learning styles describe different preferences in how learners perceive and retain information, and describe “a distinctive and habitual manner of acquiring knowledge, skills or attitudes through study or experience” [Popescu 2010]. Kozhevnikov, Evans and Kosslyn [2014] suggest to consider all different models about cognitive style as patterns of adaptation to the environment. Such models describe cognitive style with various attributes, such as analytic approach, learning structure, group orientation, instructional preference, etc. Liu and Ginther [1999] cluster cognitive style models into field dependence vs. field independence (in regard to the ability to distinguish key elements from a distracting or confusing background), holistic vs. analytic (in regard to the approach of facing problem complexity), hemispheric preferences (in regard to logic thinking and creativity), as well as sensory preference (in regard to the preferred way of information presentation and processing). Coffield et al. [2004] provide an overview and review of 13 major models of learning styles.

[1] Even though this thesis has its focus on perception in regard to sensory systems and one would thus perhaps first think of the term “sensory preference”, we favor the term “perceptual preference” for two reasons: on the one hand, we do not want to exclude other perceptual systems that are not directly linked to a sensory system (which is, e.g., the case for several learning strategies models), and on the other hand the term “sensory preference” has a special notion as it is mostly used in the context of food and taste research whereas “perceptual preference” is wider in scope.

Generally, a broader spectrum of approaches to information processing and learning allows the learner a higher flexibility and, if necessary, an adaptation of the individual approach to the respective learning content or task. Still, perhaps by habit, perhaps due to insufficient meta-knowledge or lack of training, most people show tendencies to rather use a limited set of cognitive approaches (which can again be seen as preferences) for every situation instead of always choosing the approach the most appropriate to the respective situation.

Most research describes cognitive styles with the help of multidimensional models. Many models combine two attributes each to bipolar dimensions, e.g. analytic vs. holistic or vs. intuitive or vs. trial-and-error, and calculate for each dimension the test taker's position on the axis between the two given extremes. Despite the fact that these bipolar models are quite common and some of them even are amongst the most cited and used ones, such bipolar descriptions are always biased as they are not able to capture the flexible use of so-called opposed attributes as described above. Therefore, for this thesis, we decide to use a description model that uses one dimension for each attribute, without anticipating that some of those attributes may only occur opposed to others. For the description and comparison of perceptual preference, we suggest the use of a more-dimensional vector. As there is no agreement in literature on the number of sensory preferences that need to be taken into account (one dominant, a combination of one, two or even more dominant systems), we opt for a structure that allows using all information available for latter calculation, which is the case for vectors.

The concept of sensory preference can be traced back to Galton [1883]. Since then, quite some models have one or more attributes that express sensory preferences: Several learning style models describe dimensions of information representation that can be classified as sensory preference for visual input as opposed to verbal input (which might be considered as preference for auditory input if one assumes content is communicated orally): e.g., the dimension "imager vs. verbalizer" as proposed by Riding [Sadler-Smith and Riding 1999; Coffield et al. 2004], or the dimension "visual vs. auditory", which was later changed to "visual vs. verbal", as proposed by Felder and Silverman [Felder and Silverman 1988; Felder 2002]. Dunn and Dunn's model includes the "modality preferences" visual, auditory, kinesthetic and tactile preference [Dunn 1988; Coffield et al. 2004].

An application of this concept to the field of language learning was published by Reid [1987]. In her model on learning styles she differentiates between four basic perceptual learning channels, visual, auditory, kinesthetic and tactile modalities, and presents a questionnaire focusing on the learner's perceptual preference and his or her preference for individual vs. group learning.

According to Dunn [1988], students achieve significantly better learning results when the students' perceptual preference is matched with educational methods. Those findings have recently been confirmed for visual and verbal information by Koć-Januchta et al. [2017]. As this concerns mainly the processes of receiving and processing information – and therefore processes that we would like to evoke and ease for our users – it might make sense to use this concept in a broader context.

2.2 Relevant Contexts: User Modeling, Learning, e-Commerce

In a next step, we take a look at contexts where perceptual preference plays a role, namely user modeling, learning, and e-commerce.

2.2.1 User Modeling

User modeling is an interdisciplinary research field, combining computer science and psychology. It aims to produce and use models of individual users or user groups in order to allow personalized interactions. Personalization, in this context, means, that content, interaction patterns, or design are adopted to or chosen according to the user's preferences. Applications of user modeling research concern various application fields, e.g. e-learning, e-commerce, targeted advertisement, or recommender systems.

In the field of user modeling, one of the challenges is thus to collect a maximum of information on the users with a minimum of direct interaction. Besides the conventional information elicitation via questionnaires, implicit knowledge sources can be, e.g., purchase data, ratings, user factual data, transactional data such as behavior, item factual data, etc. [Brun, Boyer and Razmerita 2010].

In the field of user modeling, two sorts of variables are modeled in order to describe a user: rather permanent and rather transitory attributes. Permanent attributes stay the same over a long time and can concern user personality, preferences, or fields of interest. Transitory attributes are state-dependent, may change after a short period of time, and represent moods, plans and goals. The choice and number of variables used in such a model are strongly dependent on the field of application and the purpose of the user model. User models are usually designed for a specific goal, e.g., to make people learn more easily or buy more. One method to evoke such effects is to personalize the choice, amount and presentation mode of information.

Two subfields of user modeling should be considered within this context: The subfield of cognitive user modeling combines psychological theories and representations about aspects of human cognition. Cognitive architectures take psychological theories (decision making, information processing) as well as test based results on human cognition into account. Such architectures structure cognitive user attributes and use them to explain and to predict user behavior. Different applications for this field are described by Heinath et al. [2007]. The subfield focusing on educational application seeks to collect information about learner characteristics like knowledge, skills, and personality traits in order to select the best learning environment for a particular student to optimize learning outcome [Shute and Towle 2003]. We suggest considering perceptual preference as a new variable for the field of cognitive styles which combines aspects of cognitive user modeling and user modeling in e-learning environments. We think that this variable could be a useful extension for the subfields of cognitive user modeling and e-learning.

2.2.2 Learning

Learning means the process of storing knowledge or methods in the brain. This can be achieved via experience, repetition, and/or understanding. But again, information as well as experience are biased and influenced by different filters such as perception and awareness, as described in section 2.1.1. Therefore, also learning is a process that depends on the learner's personality, preferences and cognitive style. "Learning is related to a number of individual cognitive and affective trait and statelike characteristics, which account for the corresponding variability in learning performance." [Tsianos et al. 2008] Learning strategies can thus be categorized as rather permanent attributes. They are developed according to conscious or unconscious interventions by the learner. Furthermore, interventions might be formed by some other external agent like teachers, tutors or learning systems [Robotham 1999]. Learning strategies might change after a longer period of time or after intense training, but not from one day to the other [Reid 1987].

Learning processes supported by computers or the internet are commonly known as electronic learning or e-learning. According to Clark and Mayer [2016], the following specific advantages of e-learning can be identified among others: reduced overall cost, increased retention, on-demand availability, self-pacing, interactivity etc. Whereas traditional teaching approaches based on classroom instruction or lecturing provide little or no options for individualization, the setting of e-learning allows customized interaction with the learner, including personalized teaching material and teaching methods. Therefore, a lot of research has been investigating the use of computers for learning within the last decades [Khan and Ally 2015]. The field of user modeling considers personalized learning as one important application field for its research results, too.

In regard to adaptation to learning styles, there are two steps that need to be performed: first the elicitation of the user's learning styles, and second the application of that knowledge to the adapted process of information presentation.

In the field of e-learning, several methods are used to obtain information on the user's learning style: besides the conventional approach via questionnaire, the user's preferred learning style can be elicited in manifold ways. Özpolat and Akar [2009] propose to consider the user's interests as an indicator for the user's learning style based on the dimensions of Felder and Silverman [1988]. Atman, Inceoğlu and Aslan [2009] build their learning styles diagnosis upon Felder and Silverman [1988], too, but they use learner behaviors as indication for learning style classification. Additionally, the learner's behavior in web based learning can be used as an indicator to detect learning styles [Kurilovas, Kubilinskiene and Dagiene 2014]. On top of learning behavioral features, also learning context features can be taken into account [Chang et al. 2009; Chi et al. 2010]. The adequate model can also be identified via recommender systems [Bourkoukou, El Bachari and El Adnani 2016].

Again, Personalization offers a huge variety of options. The effects of addressing learning styles have been subject of controversial debate [Cuevas 2015; Truong 2016]. Tsianos et al. [2010] report a positive effect on academic performance when taking the users' working memory span into account. Chi et al. [2010] apply Reinforcement Learning in an educational dialogue system and suggest adding domain-oriented and system behavior related features as effective pedagogical strategies in natural language tutoring systems. Parvez and Blank [2008] provide a feedback infrastructure based on the learning style model dimensions of Felder and Silverman [1988]. According to their research, students show better learning effects with learning style based feedback. Huang, Hwang and Chen [2016] show that matched learning style goes along with a physical reaction caused by positive emotion and leads to better learning performance.

2.2.3 E-Commerce

The Oxford dictionary describes electronic commerce as "commercial transactions conducted electronically on the Internet" [Oxford Dictionaries 2010]. As a research field, electronic commerce has a more general meaning and deals with all topics in the context of commerce. Therefore, it is highly interdisciplinary and integrates research from business and management, computer and information science, industrial engineering and operation research, engineering, economics, law, and others [Wang and Chen 2010].

Within our scope, one of the most interesting subfields deals with consumer preference. Consumer preference refers to an individual's, consumer group's or company's preference for one product out of a range of comparable offers. Research on this topic focuses on the structure of consumer preference itself, on influence factors on consumer preference as well as on marketing and other strategies to bind consumer preference to certain trademarks or products [Chung and Rao 2012]. In that concern, user modeling and e-commerce research are an effective combination.

2.3 Existing Test Methods

User preference for an item can be determined by a number of attributes of the item and the user's preferences in regard to those attributes. Such complex preference structures can be represented with multi-attribute decision systems. Clustering the attributes to sets allows induction from some information to an overall picture of the user's preferences. There are two starting points for such induction processes, namely content similarity (the user liked item *x*, *x* is similar to *y*, so the user might also like *y*) and user similarity (user *A* has a high similarity to user *B*, user *B* likes *x*, so user *A* might also like *x*), which are fundamental for recommender systems [Ricci, Rokach and Shapira 2015].

Even though it seems obvious that sensations are graded in strength and hence should be expressed in a scalable manner, measurement of sensation is crucial. One challenge lies in the fact that sensations cannot or only partly be observed directly, but only based on reports. Already for external sensation, it is a challenge to define adequate methods of measurement, but that is even harder when it gets to internal sensation [Laming 1997]. Judgments of sensation are influenced by psychological judgment, physical intensity and duration of the sensation [Kornbrot 2016], and can be addressed with the help of sensory evaluation practices [Kemp, Hollowood and Hort 2009; Stone, Bleibaum and Thomas 2012] like, e.g., perceptual maps [Ferreira Santos, Rodrigues Liska and Angelo Cirrillo 2017].

Another challenge of preference elicitation lies in the fact that many expressions of preference are only present on a subconscious level and are thus constructed at the very moment when people are asked a valuation question [Payne, Bettman and Schkade 1999], even though preference-influenced behavior is rather coherent [Barkan, Ayal and Ariely 2016].

In order to acquire information concerning user preferences, one can choose out of a huge variety of methods. Among them, there are mainly two approaches to be distinguished: explicit and implicit testing.

Explicit testing is the classic approach to the acquisition of information. It works straight-forward and uses direct questions addressing the field of interest. For this approach, a dedicated measuring instrument, e.g. a psychological questionnaire, is needed. It leads to a static learner model.

Implicit testing techniques use existing information and interpret them in order to get information about the field of interest. In the field of user modeling, implicit testing means to unobtrusively obtain information about users by analyzing their behavioral patterns taken from natural interactions with the system and to infer information out of existing data like self-defined user profiles or blog text. The primary advantage of implicit techniques is that such techniques remove the cost to the user of direct interactions with the system by filling in questionnaires or providing feedback, etc. Furthermore, those methods may be more accurate than traditional ones, and can lead to a dynamic learner model, which can be regularly updated during the learning process [Popescu 2010].

Even though implicit measures are generally thought to be less accurate than explicit measures, there are some experiments that report comparable results for both methods [White, Ruthven and Jose 2002]. Furthermore, implicit methods are attractive as they allow gathering large quantities of implicit data at no extra cost to the user. Moreover, implicit measures can be combined with explicit ratings to obtain a more accurate representation of user interests. Sources of implicit feedback on user preferences include reading time, saving, printing and selecting [Gawronski and De Houwer 2014].

Sentiment detection targets to extract the author’s sentiments towards the discussed topic(s) out of text. The research field of sentiment detection deals with similar problems and challenges as it also tries to infer structured data from unstructured information, and can hence be considered as a basis for the research presented in this thesis. Information can be gained on a lexical (word level). An overview of methods and algorithms for emotion mining and sentiment detection is given by [Medhat, Hassan and Korashy 2014; Yadollahi, Shahraki and Zaiane 2017].

CHAPTER 3

PPQ: Perceptual Preference Questionnaire[2]

The world is full of magic things, patiently waiting for our senses to grow sharper.
(William Butler Yeats)

The first test approach developed within this thesis is an explicit testing method. The PPQ is a newly developed questionnaire to assess aspects of perceptual preferences in regard to information processing, knowledge gain, and learning. Even though there exist some tests on cognitive style mentioned in section 2.1 that partially test sensory preferences, to the best of our knowledge there is none that meets our needs, which are (a) a focus on perceptual preference only, (b) a relation to individual information processing and learning without testing other aspects of learning styles, and (c) to be in German language so that it can be understood by our survey participants. The PPQ is designed to meet those requirements. It is validated with n=76 participants, whereof 52 male and 24 female, with a mean age of 37.24 years (*SD* = 13.28).

In a second step, we examine co-occurrence patterns of perceptual preference and interests. As, due to reliability issues, we apply the method of explicit testing concerning perceptual preference, we decided to elicit information about the users' interests analogously. Nonetheless, this step could be replaced by implicit testing or by using existing knowledge about the user if available.

[2] This chapter is based on the publication „I-Know my Users: User-Centric Profiling Based on Interests and Perceptual Preferences“ [Kellner and Berthold 2012].

3.1 PPQ: Test Design

The PPQ is composed of four scales, one for each sensory system: visual, auditory, kinesthetic, and olfactory and gustatory (widely addressed as “the five senses”). Olfactory and gustatory are treated as one scale because of the high overlap within those two fields. This overlap might be explained by the similar body regions that are active when perceiving odor and taste, wherefore they are often jointly referred as “chemical senses” [Goldstein and Brockmole 2016].

The design of the test items was built upon an extensive study of literature on perceptual preference. The PPQ’s test items are inspired by related work (see chapter 2) as well as by the test items of Reid’s Perceptual Learning Style Preference Questionnaire (PLSPQ) [Reid 1987] and some non-validated online self-assessment tools such as Chislet/Chapman’s VAK test, Fleming’s VARK test, and its German equivalent, Stangl’s HALB test.[3]

On a 5-point-Likert scale, participants can indicate their agreement with each test item on answer options ranging from “strongly disagree” to “strongly agree”.

With regard to item validation, we developed 46 items in order to choose the most relevant items per category based on validity tests. For those 46 items, we checked all specific values considering test difficulty, test reliability and structural validity by factor analysis [Bühner 2011], and identified the items that provide best values regarding test construction without losing information. The 46 test items candidates for the PPQ can be found in Table 1 (the complete list is given in German only, the chosen items are provided with an English translation, see Table 2 and Table 3).

For statistical analyses, we used IBM SPSS, Version 19.0. If not explicitly mentioned, statistical requirements for inference statistical analyses and procedures were fulfilled. For all analyses, alpha level was α = .05.

[3] More information on those online assessment tools can be found at the respective sites, VAK Test http://www.businessballs.com/vaklearningstylestest.htm, VARK Test http://www.vark-learn.com, and HALB Test http://arbeitsblaetter.stangl-taller.at/TEST/HALB/Test.shtml.

Expected Scale	Item Text
Visual	Ich kann mir Fakten besser merken, wenn ich sie lese.
	Einen Text zu lesen reicht mir nicht aus, damit ich mir die Inhalte merken kann.
	Ich bevorzuge Hörbücher gegenüber dem Lesen von Büchern.
	Ich verstehe Anleitungen nur, wenn ich sie selber lese.
	Wenn Freunde auf Urlaub waren, finde ich deren Urlaubsbilder interessanter als ihre Erzählungen.
	Zur Orientierung in einer fremden Stadt verwende ich am liebsten einen Stadtplan.
	Ich kann mir die meisten Sachen nur merken, wenn ich sie auch gelesen habe.
	Namen kann ich mir viel besser merken, wenn ich sie lese, als wenn ich sie höre.
Auditory	Beim Lernen spreche ich mir die Lerninhalte selbst vor, damit ich auch höre, was ich lese.
	Ich lasse den Fernseher gern im Hintergrund laufen, auch wenn ich das Bild nicht sehen kann.
	Ich spiele aufgenommene Lerninhalte mehrmals über ein Audiogerät ab.
	Zuhören reicht mir, damit ich mir Dinge gut merke.
	Ich telefoniere lieber, als dass ich E-mails lese bzw. schreibe.
	Wenn mir jemand etwas erklärt, komme ich meist besser zurecht, als wenn ich nur einen Text mit Erklärung lese.
	Um etwas in einem unbekannten Ort zu finden, frage ich am liebsten Einheimische um Hilfe, die mir den Weg erklären.
	Erklärungen reichen mir meistens nicht aus, ich benötige etwas Schriftliches, um etwas gut zu verstehen.
	Hörspiele finde ich besser als Stummfilme.
	Ich lasse gern nebenbei Musik bzw. den Fernseher laufen, um ein Hintergrundgeräusch zu haben.
	Wenn ich eine Kunstausstellung besuche, lasse ich mich sehr gern durch den Audio Guide informieren
	Ich gehe lieber zu einem klassischen oder Jazzkonzert, als in eine Kunstausstellung.
Kinesthetic	Ich bevorzuge es, mit meinen Händen zu arbeiten.
	Lernen kann ich am besten, wenn ich auch Gegenstände in die Hand nehmen kann.
	Handwerkliche Tätigkeiten sind für mich sehr wichtig.
	Wenn ich bestimmte Bewegungen sehe, möchte ich sie am liebesten gleich selbst ausprobieren.
	Erzählungen bringen mir nicht so viel, ich möchte die Sachen lieber selbst erleben.
	Bei der praktischen Umsetzung lerne ich viel mehr als beim Lesen.
	Ich bevorzuge "learning by doing" und probiere alles lieber selbst gleich, statt erst aus Büchern zu lernen.
	Wenn ich eine neue Frucht entdecke, möchte ich sie auf jeden Fall gleich anfassen.
	Wenn ich viel lerne, brauche ich viel Bewegung.
	Ich muss nicht alles anfassen, was ich gern mag.
	Wenn ich selbst ein Modell baue, kann ich mir die Sachen leichter merken.
	Wenn ich etwas nicht gleich verstehe, reagiere ich emotional.
	Neue Lerninhalte wecken bei mir häufig Gefühle.
	Meinen Gefühlen schenke ich im Alltag kaum Beachtung
	Wenn ich glaube, dass ich etwas nicht kann, bemerke ich Veränderungen in meinem Körper.
	Es ist mir wichtig, beim Lernen körperlich entspannt zu sein.
	Ich setze meine Lernpausen dann, wenn mein Körper mir das Gefühl gibt, dass ein Pause nötig ist.
	Ich achte häufig auf die Signale, die mir mein Körper gibt.
Olfactory + Gustatory	Durch bestimmte Gerüche erinnere ich mich sofort an Ereignisse oder Orte aus meiner Vergangenheit.
	Ich kann mir Gegenstände viel besser merken, wenn ich auch deren Geruch kenne.
	Mein Geruchssinn ist mir sehr wichtig.
	Dem Geruch von Speisen oder Personen schenke ich kaum Beachtung.
	Beim Lernen verknüpfe ich gern Inhalte mit verschiedenen Gerüchen.
	Bei einer neuen Frucht sind mir Geschmack und Geruch am wichtigsten.
	Rezepte kann ich mir gut merken, wenn ich den Geschmack der Zutaten dazu kenne.
	Ich probiere sehr gern verschiedene Geschmacksrichtungen aus.
	Mit Gerüchen verbinde ich verschiedene Lerninhalte.

TABLE 1 THE 46 PPQ TEST ITEM CANDIDATES

With a total of 20 remaining items, the PPQ provides sufficient information and is more economic in regard to survey length. The 20 PPQ items can be found in Table 2.

Nr.	Item Text	Scale
Vis1	Ich kann mir Fakten besser merken, wenn ich sie lese.	Visual
Vis2	Ich verstehe Anleitungen nur, wenn ich sie selber lese.	
Vis3	Zur Orientierung in einer fremden Stadt verwende ich am liebsten einen Stadtplan.	
Vis4	Ich kann mir die meisten Sachen nur merken, wenn ich sie auch gelesen habe.	
Vis5	Namen kann ich mir viel besser merken, wenn ich sie lese, als wenn ich sie höre.	
Aud1	Erklärungen reichen mir meistens nicht aus, ich benötige etwas Schriftliches, um etwas gut zu verstehen.	Auditory
Aud2	Beim Lernen spreche ich mir die Lerninhalte selbst vor, damit ich auch höre, was ich lese.	
Aud3	Hörspiele finde ich besser als Stummfilme.	
Aud4	Ich lasse gern nebenbei Musik bzw. den Fernseher laufen, um ein Hintergrundgeräusch zu haben.	
Aud5	Ich gehe lieber zu einem klassischen oder Jazzkonzert als in eine Kunstausstellung.	
Kin1	Ich bevorzuge es, mit meinen Händen zu arbeiten.	Kinesthetic
Kin2	Lernen kann ich am besten, wenn ich auch Gegenstände in die Hand nehmen kann.	
Kin3	Handwerkliche Tätigkeiten sind für mich sehr wichtig.	
Kin4	Ich bevorzuge "learning by doing" und probiere alles lieber selbst gleich, statt erst aus Büchern zu lernen.	
Kin5	Wenn ich selbst ein Modell baue, kann ich mir die Sachen leichter merken.	
Olg1	Durch bestimmte Gerüche erinnere ich mich sofort an Ereignisse oder Orte aus meiner Vergangenheit.	Olfactory + Gustatory
Olg2	Mein Geruchssinn ist mir sehr wichtig.	
Olg3	Beim Lernen verknüpfe ich gern Inhalte mit verschiedenen Gerüchen.	
Olg4	Rezepte kann ich mir gut merken, wenn ich den Geschmack der Zutaten dazu kenne.	
Olg5	Mit Gerüchen verbinde ich verschiedene Lerninhalte.	

TABLE 2 PPQ TEST ITEMS

For better traceability, an English translation of the items (which is not validated yet and would require additional testing before use) is provided in Table 3.

Nr.	Item Text	Scale
Vis1	I retain facts better if I read them.	Visual
Vis2	I only understand instructions if I read them myself.	
Vis3	For orientation in a foreign city, I prefer to use a city map.	
Vis4	I usually remember content only if I read it.	
Vis5	I retain names much better if I read them than if I hear them.	
Aud1	Reading a text is not enough for me to retain content.	Auditory
Aud2	I usually talk the learning content through to hear what I am reading.	
Aud3	I like radio plays better than silent movies.	
Aud4	I like to have music or the TV turned on to have some sound in the background.	
Aud5	I prefer going to a classical or jazz concert rather than going to an art exhibition.	
Kin1	I prefer to work with my hands.	Kinesthetic
Kin2	I learn best if I can touch objects with my hands.	
Kin3	Craft activities are really important for me.	
Kin4	I prefer "learning by doing" instead of learning from books.	
Kin5	I retain content easier if I build models.	
Olg1	Certain odors immediately remind me of special events or places from my past.	Olfactory + Gustatory
Olg2	My sense of smell is really important for me.	
Olg3	When studying, I like to connect content to different odors.	
Olg4	I retain recipes well if I know the taste of the ingredients.	
Olg5	For me, different odors are related to different learning content.	

TABLE 3 PPQ TEST ITEMS: NOT VALIDATED ENGLISH TRANSLATION

3.2 Methodology

3.2.1 The Online Survey

The online survey serves mainly two purposes of data collection. First, data was collected to analyze and validate the PPQ, and second, interests, forum membership and the actual scores of the PPQ were collected. The survey is composed of five blocks of content. It takes around ten minutes to answer the questions. The five blocks of content are:

- **Introduction and explanations:** The user gets some information on the survey topic and an example on how to fill in the questionnaire.
- **Demographic data:** Some basic demographic data is collected, namely gender, age, highest education level, and in which forum(s) the user is active.
- **Interests:** In order to ensure a high reliability of results, we opted for explicit elicitation of interests. Following the examples of GMX[4] and YahooGroups[5], we composed a list of 27 fields of interests that are shown to the user in random order (see Table 4). The user is first asked to choose a free number, but at least five elements out of that list. In a second step, he or she is asked to put the five most important interests into a ranking list.
- **PPQ:** The Perceptual Preference Questionnaire (PPQ) was tested in its initial version with 47 items. After a detailed analysis of results and item validation, we suggest to use a shortened version with 20 items only for further research.
- **Prize draw:** In order to raise the participation level, we combined the online survey with the option to win a prize by filling out our questionnaire.

art	handicrafts and do-it-yourself	politics
cars	languages	psychology
computers	literature	religion
cooking and baking	look and style	self-awareness
eating	movies	society
economy	music	sports
eroticism	nature	theater
finances	philosophy	travelling
games	photography	wellness

TABLE 4 LIST OF 27 INTERESTS

[4] http://www.gmx.net/
[5] http://de.groups.yahoo.com/

The online survey as well as the explanations are in German language.

3.2.2 Procedure

We used the Unipark[6] survey system to set up the online survey as described above. The call for participation was posted in four discussion forums. The forums were chosen by the following three criteria: (1) They should have a large and active community (determined by the number of users and the number of posts), (2) be German-speaking (in order to understand our questionnaire), and (3) should deal with a topic, which is expected to be thematically linked to one sensory preference. The topics of the chosen forums are photography (one German and one Austrian forum), music (one German forum), and cooking (one German forum). By addressing users from topically not related forums, we assure a high probability that all preference types are represented in the evaluation sample in order to conduct analyses of the questionnaire items and its structure as a whole.

The first call for participation was the same in all four discussion forums. If requested, we gave more information on the research process itself and its goals, individually answering the questions raised by forum users of the respective forum. The need for information differed a lot among the forums.

3.2.3 Data Description

We only used completely filled-in questionnaires for further analyses, which leads to a number of n=76 participants, whereof 52 male and 24 female, with a mean age of 37.24 years (SD = 13.28). The educational background of the sample was allocated by 2 without school-leaving qualifications, 15 with a secondary school degree, 5 with apprenticeship, 28 with high-school diploma, 22 with university degree, and 4 with a post gradual degree. The distribution of forum membership in the test sample is as follows: cooking forums (13), photography forums (23), and music forums (37). Data that could not be linked to a forum was not counted in for analyses in regard to forum membership.

[6] http://www.unipark.com/

3.3 Test Validation

A three step validity analysis is applied to the selected 20 items.

In a first step, descriptive statistics and item difficulty (p_i)[7] were analyzed (see Figure 7 and Table 5). The 20 relevant items provide acceptable means (*M*) and standard deviations (*SD*) and lie between the critical p_i limits of .20 to .80 defined in Bortz and Döring [2006]. Means of the PPQ per scale are: visual (3.49; *SD* = .65) auditory (*M* = 3.18, *SD* = .68), kinesthetic (*M* = 3.52, *SD* = .80) and olfactory (*M* = 2.93, *SD* = .83).

In a second step, factor analysis was performed. Extraction of the number of factors was done by the *Kaiser*'s eigenvalue-greater-than-one-rule and the scree-test [Cohen 1988], which identified six factors. However, the *Kaiser* criterion overestimates the number of factors in most of the cases [Zwick and Velicer 1986]. Four factors seem to fit the data, which is confirmed by the major drop in the scree plot (see Figure 6).

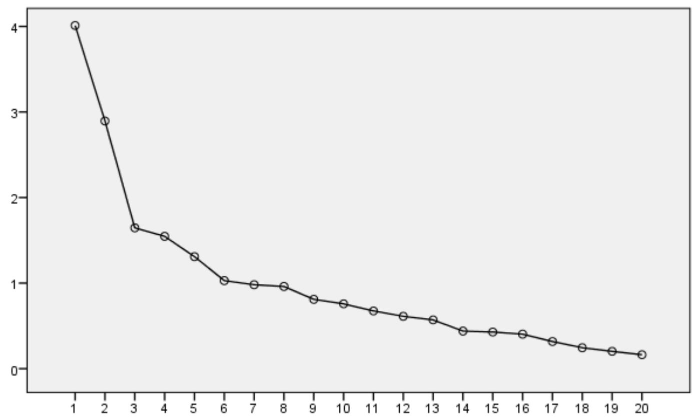

FIGURE 6 FACTOR ANALYSIS SCREE PLOT: EIGENWERT (X-SCALE) PER NUMBER OF FACTORS (Y-SCALE)

[7] Note: *n* = 76; *Min* =1 and *Max* =5 for all 20 items; k is the number of answer options on the rating scale, in this case *k*=5. p_i index was calculated by the following formula for rating scales that start the rating by 1:

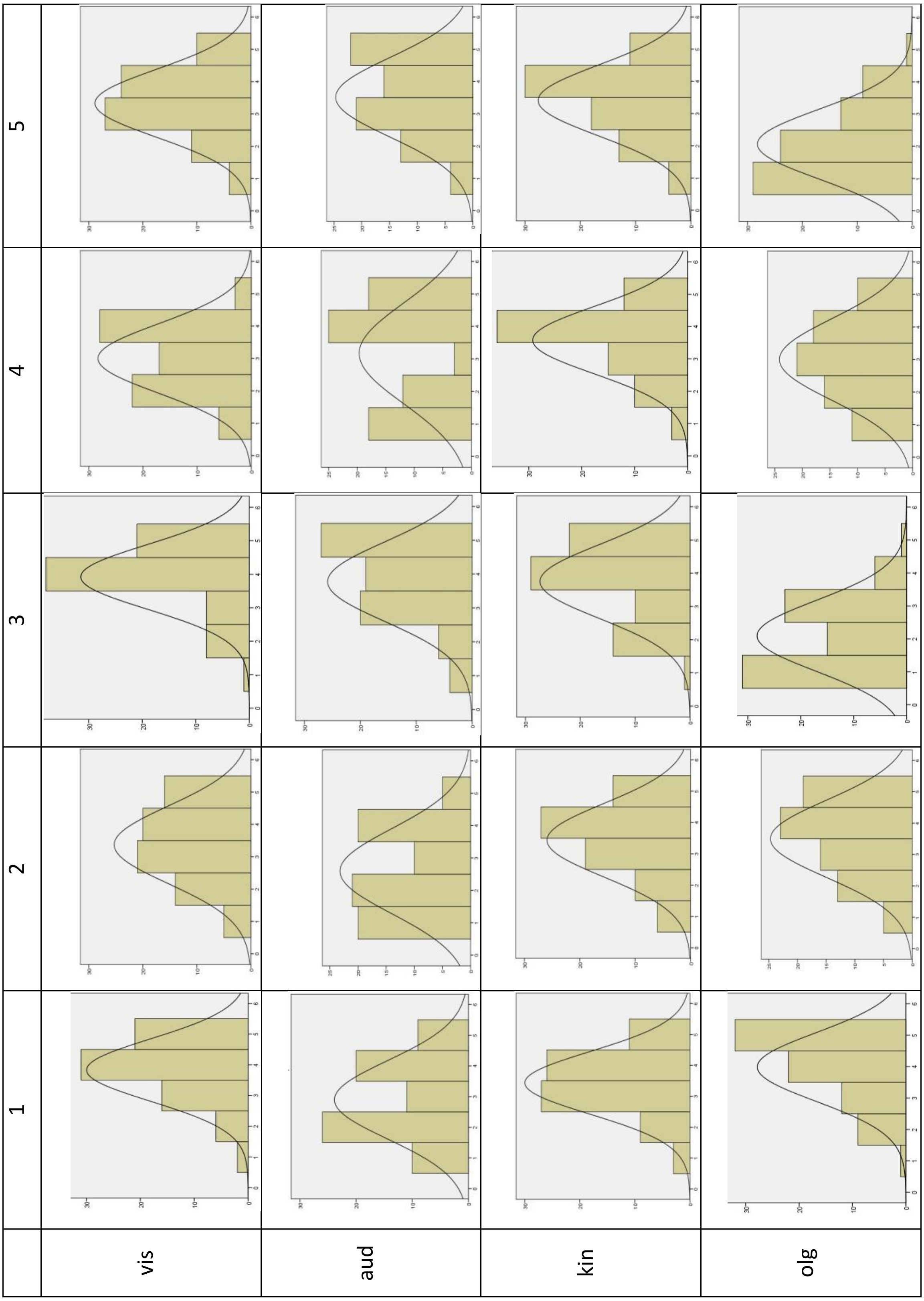

FIGURE 7 PPQ TEST ITEMS: FREQUENCY (X-SCALE) PER LIKERT-SCALE OPTION 1-5 (Y-SCALE)

An additional parallel analysis, as proposed by Horn [1965], also suggests four factors, which account for 50.49 % of the total variance. After a Varimax rotation, factor 1 (olfactory+gustatory) accounts for 16.32%, factor 2 (kinesthetic) for 15.29%, factor 3 (visual) for 10.20% and factor 4 (auditory) accounts for 8.68% of the variance. An orthogonal Varimax rotation was applied in order to obtain the factor structure of the PPQ, which shows that each item loads highest for the according factor with smaller or negative secondary loadings for other factors (see Table 5). Item *Aud2* is an exception with a higher loading on factor 1, but from context it fits much better for factor 4, which is why it has been assigned to factor 4.

In a third step, reliability was analyzed based on *Cronbach*'s alpha (α_C) for each scale. α_C represents the internal consistency of the items within a scale; additionally, α_Cs are reported if a specific item would be deleted (see Table 5). *Vis3* is the only item that would increase α_C if deleted, but the rise would not change alpha dramatically. For reasons of well-balanced scales, *Vis3* is included in further analyses. The scales of visual and auditory perception show rather low α_C and might need revision.

The PPQ's validation results for the three validity test steps, including (1) mean (M), standard deviation (SD), and item difficulty (pi), (2) factor analysis for the four factors F1-F4, and (3) reliability testing based on Cronbach's alpha (αC) and Cronbach's alpha if item deleted (αC iid), can be found in Table 5.

		Step 1: Descriptive Data			*Step 2: Factor Analysis*				*Step 3: Reliability*	
Nr.	Scale	*M*	*SD*	p_i	F1	F2	F3	F4	α_C	α_C iid
Vis1		3.83	1.01	0.71			0.38	-0.53		0.56
Vis2		3.37	1.20	0.59			0.71			0.46
Vis3	Visual	3.92	0.96	0.73	-0.36		0.36		.58	0.66
Vis4		3.00	1.07	0.50	0.29		0.66			0.42
Vis5		3.33	1.05	0.58	0.39		0.65			0.45
Aud1		2.89	1.27	0.47		0.38		0.67		0.19
Aud2		2.59	1.31	0.40	0.44			0.36		0.29
Aud3	Auditory	3.78	1.17	0.70				0.62	.33	0.22
Aud4		3.17	1.54	0.54			0.35	0.41		0.25
Aud5		3.51	1.23	0.63		-0.42		0.25		0.42
Kin1		3.43	1.01	0.61	0.28	0.78				0.72
Kin2		3.41	1.10	0.60		0.63	0.28			0.76
Kin3	Kinesthetic	3.43	1.17	0.61		0.77			.79	0.75
Kin4		3.75	1.11	0.69		0.69		0.28		0.75
Kin5		3.58	1.04	0.65	-0.28	0.65				0.75
Olg1		3.99	1.09	0.75	0.53					0.77
Olg2		3.50	1.23	0.63	0.64			-0.24		0.76
Olg3	Olfactory + Gustatory	2.09	1.07	0.27	0.84				.78	0.69
Olg4		3.00	1.25	0.50	0.66					0.75
Olg5		2.07	1.08	0.27	0.78					0.70

TABLE 5 PPQ TEST ITEMS

3.4 PPQ and Interests

Even though the Perceptual Preference Questionnaire (PPQ) is a valid instrument, we are aware that knowledge discovery research is rather looking for methods that require less intense interaction with the user than extensive questionnaires in order to design personalized applications. The PPQ is thus thought to be an interim step, which might be replaced by knowledge about the users' perceptual preferences and interests. In a next step, we compare the PPQ results with results from a forum text analysis and examine co-occurrence patterns of perceptual preference and interests.

Modeling the user's interests is one of the most central topics in the field of user modeling. Interests might be handled as rather transitory attributes in case of online shopping or short term information need, but can also be rather permanent in case of hobbies or general fields of interest [Jiang and Sha 2015]. Knowledge about the user's interests can, again, be gained in explicit and implicit manner. Explicit information can be collected by using questionnaires or surveys, which tend to be annoying to the user who does not want to spend time on answering questions when he/she opens a new website or tries out a new online service or software. Hence, a lot of research has been done on how to gain knowledge about the user's interests without needing explicit content. One method used by a lot of recommender systems is to handle past activities (like purchases) as indicators to a user's interests [Klašnja-Milićević, Ivanović and Nanopoulos 2015]. Furthermore, information about the user's interests can be accessed via analysis of browsing activity, social network interactions and partners, user-generated text like e-mails or forum posts, and queries in a search engine [Guy et al. 2013; Gasparetti 2017; Liang et al. 2017]. The other way round, interests can be seen as an indication of personality [Volkova, Bachrach and Durme 2016]. Still, reliability and validity of explicit and implicit methods might differ [Golijani-Moghaddam, Hart and Dawson 2013].

The analyses are grouped into two parts and investigate distribution and co-occurrence patterns of interests and of perceptual preference as well as possible correlations between perceptual preference and interests, perceptual preference and membership of certain forums, and interests and membership of certain forums.

Perceptual preference is abbreviated as follows: visual (V), auditory (A), kinesthetic (K), olfactory+gustatory (OG).

3.4.1 Distribution and Co-occurrence Patterns

Co-occurrence measures (A_B for 2 attributes A and B) were calculated with the formula $A \cap B/(A \cup B - A \cap B)$. Values range of $0 \leq A\_B \leq 1$, with 0 for attributes that never apply jointly, and 1 for attributes that only apply jointly.

Results for perceptual preference were split into high and low groups for each factor using median split. This resulted in 38 participants for each high/low group per sensory system (high/low V (Md = 3.40), high/low A (Md = 3.20), high/low K (Md = 3.60) and high/low OG (Md = 2.80)). Perceptual preferences occurred in the following groupings: 8 users with zero high preferences, 21 users with one and 21 with two high preferences, 15 users with three high preferences and 11 users with four high preferences. Co-occurrence of two preferences was highest for V_OG (.52), close to the mean (.41+.02) for V_K, A_K and K_OG, and lowest for A_OG (.36) and A_V (.33).

Concerning interests, we first filtered out less relevant interests. Interests were considered as relevant and included in further analysis if chosen by at least 20% of the participants of our study, which was the case for 18 interests. The most relevant ones were music (54 users), photography (41 users) and computers (36 users). Co-occurrence was calculated for all relevant interests and is highest for Cooking+Baking_Eating (.61) and lowest for interests that were never chosen jointly (0). Figure 8 shows an analysis of the 18 relevant interests as a graph. Interests are presented with respect to frequency of occurrence (size of the node) and co-occurrence with other interests (color and width of edges in four categories: from thin to thick ≥.35, >.40, >.45 >.50).

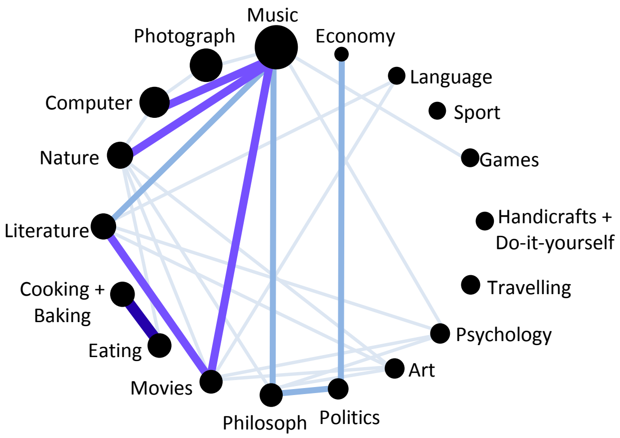

FIGURE 8 CO-OCCURRENCES OF INTERESTS

3.4.2 Correlations

Phi-correlations (for nominal data) were computed. The phi-correlation coefficient (r_φ) and significance level are presented. Note that only significant results are presented here; as suggested by [9], the level of correlation is considered to be r = .1 small correlation, r = .3 medium correlation, r = .5 high correlations.

3.4.2.1 Perceptual preference and interests

Phi-correlation analyses showed that interests in music ($r_{\varphi 76}$ = -.232, p = .04) and nature ($r_{\varphi 76}$ = -.239, p = .04) correlated negatively with visual preferences indicating that users who are interested in music or nature prefer other than visual input.

Medium to high negative correlations were observed for the interests of art ($r_{\varphi 76}$ = -.252, p = .03) and photography ($r_{\varphi 76}$ = -396, p = .00) with auditory preferences. These results show that users who are interested in art and photography prefer other than auditory input.

Interests in psychology ($r_{\varphi 76}$ = -.308, p = .01) and travelling ($r_{\varphi 76}$ = -.283, p = .01) correlate negatively with kinesthetic preferences, representing a lack of interest in psychology and travel when at the same time kinesthetic preferences are high.

No significant correlations were observed for olfactory/gustatory preferences.

An overview of the observed correlations is given in Figure 9. Green edge color stands for positive and red for negative correlation.

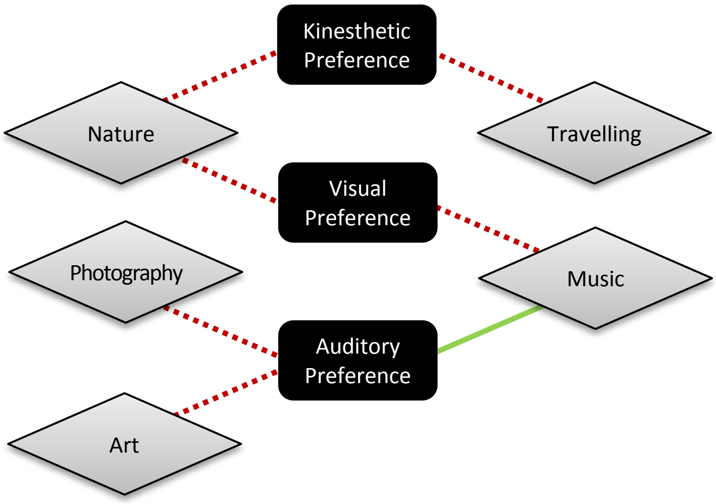

FIGURE 9 CORRELATIONS BETWEEN PERCEPTUAL PREFERENCES AND INTERESTS

3.4.2.2 Perceptual preference and membership in certain forums

In regard to perceptual preference and membership in certain forums, results show that membership of a music forum correlates with auditory preferences ($r_{\varphi 76}$ = .342, p = .00), stating that music forum members prefer auditory input. At the same time, membership in a photography forum correlates negatively with auditory preferences ($r_{\varphi 76}$ = -.441, p = .00), indicating that members of photography forums prefer other than auditory input.

Membership in cooking forums correlates positively with olfactory/gustatory preference ($r_{\varphi 76}$ = .245, p = .03). This means members of cooking forums tend to have a preference for olfactory/gustatory perception.

3.4.2.3 Interests and membership in certain forums

Investigating the correlations of membership in certain forums and interests, we found highly positive correlations between membership in a music forum and the interest for music ($r_{\varphi 76}$ = .622, p = .00), which underlines that membership in a music forum is most likely accompanied by interest in music. Furthermore, also the interest in computers correlates positively with a membership in a music forum ($r_{\varphi 76}$ = 237, p = .04) suggesting that people who are active in music forums tend to be interested in computers. At the same time, interest in music correlates negatively with cooking forum ($r_{\varphi 76}$ = -.326, p = .00) and photography forum ($r_{\varphi 76}$ = -.464, p = .00) participation. This shows that members in cooking or photography forums have low interest in music.

Cooking and baking correlates positively with a membership in a cooking forum ($r_{\varphi 76}$ = .476, p = .00) and negatively with a membership in a music forum ($r_{\varphi 76}$ = -.273, p = .02). These results indicate that users interested in cooking and baking join cooking forums, but do rather not join music forums.

Interest in photography is strongly correlated with a membership in a photography forum ($r_{\varphi 76}$ = .511, p = 00), but negatively correlated with a membership in a music forum ($r_{\varphi 76}$ = -.442, p = .00). Therefore, users who take part in the community of photographers are most likely interested in photography, whereas members of music forums are rather not interested in photography.

Interest in movies is negatively correlated with membership in a photography forum ($r_{\varphi 76}$ = -.282, p = .01), indicating that users who are active in photography forums show low interest in movies. Negative correlations were also observed between membership of a photography forum and the interest in games ($r_{\varphi 76}$ = -.247, p = .03) as well as sport and languages (for both correlations: $r_{\varphi 76}$ = -.231, p = .04), which means that members of photography forums are rather not interested in games, sport, and languages.

CHAPTER 4

The Vocabulary of Perception

All our ordinary verbal expressions bear
the stamp of our customary forms of perception.
(Niels Bohr)

The second test approach developed within this thesis is an implicit testing method. It connects psycholinguistic research with user modeling and investigates the information potential that lies in user generated text.

4.1 Introduction

According to psycholinguistic research, any text contains a lot of implicit information about its writer. The Internet provides an incredible amount of text produced by users. The information potential of text that can be directly linked to a user (which is especially the case for forum and blog posts) is not yet sufficiently examined. Natural language processing techniques are in some cases used for getting information on the user's personality, moods, affects and sentiments, but not yet for getting information on the user's cognitive style concerning perceptual preference. The following chapter of this thesis explores the potential of such an idea.

Therefore, we first give an overview of related research in the context of psycholinguistics and text mining. Section 4.3 deals with lexical expression of perception and the vocabulary of perception. Section 4.4 focuses on lexical analysis algorithms and the relation between stems or other semantic units and given words.

Section 4.5 presents the Signal Term Extraction Method (STEM) Algorithm, which was developed to identify the vocabulary of perception within text. In section 4.6, we embed that algorithm into the VAKOG Analysis Pipeline, which allows analyzing forum posts with regard to the use of sensory vocabulary. Section 4.7 investigates the use of perceptual vocabulary in forums.

4.2 Related Research

First, we give an overview of the research field of psycholinguistics. The second part shows in what manner text may be used as a source of implicit information about its author.

4.2.1 Psycholinguistics

Every text contains a message that might sometimes be easily understood, but sometimes needs more context to be interpreted in the right way. A text also reveals a great deal about its author. When investigating individual language use, it makes sense to take a look at psycholinguistic research. Psycholinguistics is a subdomain of linguistics that deals with psychological aspects of language use. Concerning the contents, it shows a high overlap to the field of language psychology, which has its roots in psychology and developed towards linguistics. Nowadays, those expressions are mostly used synonymously [Harley 2017].

The *language user model* [Dijkstra and Kempen 1993] sums up and structures the most important aspects of psycholinguistic research. It focuses on the brain activities during language processing and differentiates the following aspects: On the one hand language reception that consists of the systems of language recognition (distinguishing speech sounds), word recognition (identifying words), sentence based analysis (splitting up phrases into units of meaning), and the conceptual system (interpreting utterances). On the other hand, language production that combines the conceptual system (forming thoughts and objectives) and the systems of grammatical coding (expressing thoughts and objectives in phrases), phonological coding (finding words and choosing the appropriate flexion), and articulation (uttering words and phrases). Furthermore, there exist a lot of phonological, morphological and syntactic rules that are stored together with the complete vocabulary and all related lexical information in the mental lexicon. Syntactically and semantically correct phrases can be built with the help of the mental lexicon and the conceptual memory, which contains the existing individual knowledge of the world [Zeevat 2014].

Even though being situated in a certain environment and referring to it, speech is not only determined by the situation itself. Besides the relations between language, brain and body there exists another relation, namely between language and personality [Boyd and Pennebaker 2017]. With the reach of adulthood, all attributes of language that are related to psychological characteristics are fully developed and fairly stable [Crystal 1987]. One can hence conclude that observable personality traits are related to an individual's choice of expression and that such individual differences can be elicited via linguistic analysis [Goldberg 2012].

4.2.2 Using Text as a Source of Implicit Information on its Author

In recent years, the field of web text mining has done considerable research on what text can tell us about authors by analyzing spontaneous utterances of users in texts such as blogs and forums.

One research field related to this topic is the investigation of subjectivity [Balahur, Mihalcea and Montoyo 2014]. Chen [2008] differentiates the dimensions non-objectivity, uncertainty, vagueness, non-objective measurability, imprecision and ambiguity. Wiebe et al. [2004] show that unique words are more often subjective than expected and that unique words are valuable cues to subjectivity. To retrieve information on the user's affects, one can analyze individual expression based on linguistic inquiry such as proportional use of different part of speech, direct speech, punctuation, complete upper case words, the average sentence length, and the variety of used vocabulary based on bag of word (BOW) calculation [Zhai and Massung 2016]. Furthermore, words and terms expressing affects and emotions are collected in an affective lexicon and used as a basic instrument to determine expression of affect [Joshi, Bhattacharyya and Ahire 2017]. Information on the users' sentiments and opinions is gathered in the same manner [Liu 2015]. The user's mood might also be influenced by the topic he or she is writing about [Balog and de Rijke 2007; Thelwall and Buckley 2013]. Such analyses are also realized in a multilingual context [Boyd-Graber and Resnik 2010].

Analogously, information on the user's personality can be extracted from text. Several studies found correlations between personality and language use using the method of linguistic inquiry and word count [Lee et al. 2007; Tausczik and Pennebaker 2010; Yarkoni 2010]. Not only single words, but also of phrases can be used as an indication to personality [Kern et al. 2013]. In spoken language corpora, one can also analyze so-called paralinguistic features like initiative-taking in conversation, utterance type features such as a command, prompt, question or assertion, as well as prosodic features such as the voice's pitch, intensity, and speech rate [Schuller 2014] or filler-words [Laserna, Seih and Pennebaker 2014]. Based on lexical analysis, one can extract information on personality differences like neuroticism, extraversion, openness, agreeableness, and conscientiousness [Nowson 2006], but also on emotional status [Woo and Ahn 2015] and affect [Newell et al. 2017] out of blog texts. [Park et al. 2015] combine linguistic inquiry and lexical analysis to extract personality cues from social network text. Even for Twitter posts, which provide only extremely short pieces of texts, it is possible to explore personality factors [Plank and Hovy 2015]. Also, indication for depression and mental illness can be deduced from social media text [Guntuku et al. 2017].

Methods in this field are quite well developed: specialized algorithms have been developed to handle different sorts of text available on the internet concerning average text length and genre, such as social media [Stieglitz et al. 2014], blogs [Waila, Singh and Singh 2013], and news [Hamborg et al. 2017], giving only one example for each category. There exist multimodal fusion approaches [Poria et al. 2017] and multilingual solutions [Dashtipour et al. 2016]. Results gained by such methods have been refined by application of nearest-neighbour approaches that utilize models of user similarity [Karampiperis, Koukourikos and Stoitsis 2014] and have been combined with real-world knowledge [Tune et al. 2016] to enhance the classification results.

Perceptual preference can be regarded as another such dimension, giving cues on the individual point of view and the individual preferences. Considering perceptual preference as another specification of the personal point of view, those preferences likely to be expressed in the individual way of verbal expression.

4.3 Vocabulary of Perception

The vocabulary of perception has mostly been described in specific aspects, but not yet in holistic lexical analysis approaches. A major proportion of the literature that deals with the vocabulary of perception of the German language concentrates on one perceptual system only and, moreover, on one selected part of speech.

Only some publications take all perceptual systems into account. Hundsnurscher [1977] looks at the distribution of part of speech in the vocabulary of perception. This is one of the few works that takes different parts of speech into consideration, but it only gives a small number of examples per perceptual system. The characteristic occurrence of perceptual expressions in phrases is described by Clément [1971] as well as Falkenberg [1989]. In another article, Hundsnurscher describes the act of verbal expression of perception and proposes a classification structure for verbs of perception [Hundsnurscher 1976]. The aspect of perceptual sensation of events as well as their verbal description is analyzed by Engelberg [2000]. The semantic change of perceptual verbs in the German language has been documented by Harm [2000]. A collection of perceptual vocabulary that covers all five perceptual systems and every part of speech is collected by the author [Kellner 2010], and has been extended to a large dictionary of perception, including lexical and semantic analysis for each word [Salamon 2017].

The most detailed analyses have been written about the visual vocabulary of perception. The first attempt to a categorization of visual vocabulary was made by Weisgerber [1929]. The first detailed analysis of the German visual vocabulary was published by Bülow [1970]. She collects 386 visual verbs and verbal expressions whereof 58 are described via semantic factor analysis. A contrastive analysis of visual verbs and their collocations in English and German is presented by Roos [1975], French and German are compared by Schepping [1982]. A very detailed system for the semantic classification of visual verbs is designed by Robering [1985]. Two large vocabulary collections with several thousand entries each focus on color terms [Seufert 1955; Jones 2013].

Auditory verbs have been described and collected by Vliegen [1988]. The act of hearing is quite close to the production and perception of sounds, which is focus of the ‘Handbook of German verbs of communication’ [Harras et al. 2004].

Kinesthetic perception may be divided into three aspects: tactile perception of tangible objects from the outside world as well as body perception and emotion perception from the inner world. The German tactile vocabulary has been explored by Schmauks [2015]. Verbal references to body parts have been collected by Materynska [2012], verbal expression of pain is described by Overlach [2008]. Motion can be perceived from external viewers as well as by internal body perception; the verbs of motion are categorized by Diersch [1972] and compared to state descriptions verbs by Gerling and Orthen [1979]. There exists a variety of publications on the vocabulary of emotion, a profound introduction to that topic is given by [Schwarz-Friesel 2013].

The olfactory vocabulary is already addressed by Weisgerber [1928] with an interlingual comparison of problems with verbal odor labeling. A contrastive analysis of the olfactory verbs in German and English is published by Fricke [1996]. The poeticity of odor descriptions in perfume advertisements is addressed by Holz [2005], word creation and semantic field transfers in the description of odors are explored by Zimmer [2006].

Gustatory vocabulary has mostly been analyzed within the research project "semantics of sensation". Two publications from this project look at flavor adjectives in advertisements [Nawrocki 2004; Buckenhüskes, Heusinger and Nawrocki 2005], another one combines the olfactory and the gustatory dimension and presents the vocabulary of food sensation [Wäsch 2005].

4.3.1 Our Vocabulary of Perception

According to [Benamara et al. 2007; Taboada 2016], better results for text based sentiment analysis can be achieved when taking not only one part of speech into account, but several ones. As it is likely that this also applies for the vocabulary of perception, we decided to take all parts of speech into consideration when building our vocabulary of perception. But since German is a compounding language and hence highly productive, it would go beyond the scope of this thesis to go through a complete dictionary in order to identify every term that shows a relation to a perceptual system. Therefore, we opt for a stem based approach that identifies word realizations based on perceptual stems. To give an example of how many terms can be built upon one stem, Table 6 shows a number of words that are all related to the stem *seh* [see] and its realizations.

We created our lexicon of sensory vocabulary as follows: It is based on the list of stems of sensory vocabulary collected by the author [Kellner 2010]. We expanded this list by incorporating a list of color terms and performed some other minor changes in accordance to the vocabulary collections presented above. We worked with four categories of sensory vocabulary: visual, auditory, kinesthetic and olfactory/gustatory. Olfactory and gustatory expressions were combined because of the high overlap of words within those two fields [Wäsch 2005; Kellner 2010]. This overlap might be explained by the similar body regions that are active when perceiving odor and taste, wherefore they are often referred jointly as "chemical senses" [Goldstein and Brockmole 2016].

Stem	Some Word Realizations
seh	Ansehen, Einsehen, sehen, seht, Sehbehinderte, sehenswürdig, Sehenswürdigkeit, Sehfehler, Sehhilfe, Sehkraft, Sehnerv, Sehschärfe, Sehschwäche, Sehtest, Sehvermögen, unübersehbar, versehen, vorhersehen, wegsehen
sieh	siehe, siehst, sieht
sah	sah, sahen, sahst
säh	sähe, sähen
sicht	Absicht, Ansicht, Einsicht, offensichtlich, Sicht, sichten, Sichtfenster, sichtlich, Sichtung, Sichtverhältnisse, Sichtvermerk, Sichtweise, Sichtweite

TABLE 6 SOME WORDS BASED ON THE STEM *SEH* [SEE] AND ITS REALIZATIONS

Hence, the lexicon of sensory vocabulary may be divided into 4 disjoint sets, namely the lexicon of visual vocabulary L_V, of auditory vocabulary L_A, of kinesthetic vocabulary L_K, and of olfactory and gustatory vocabulary $L_{[OG]}$. The cardinalities of these sets are $|L_V|=87$, $|L_A|=92$, $|L_K|=112$, $|L_{[OG]}|=65$. Naturally, this list does not yet cover all existing terms in the German language, still, there is a huge number of word realizations that can be built with those stems as can be seen in section 4.7.

4.4 Lexical Analysis Algorithms[8]

It is not sufficient to know the vocabulary of perception and its stems, but it is also necessary to identify it within text. Text mining on a lexical basis is quite well developed for the English language. In compounding languages, however, lexicalized words are often a combination of two or more semantic units. New words can easily be built by concatenating existing ones, without putting any white spaces in between.

[8] This section is based on the publication „Algorithms for the Verification of the Semantic Relation Between a Word and a Given Word Fragment" [Kellner and Grünauer 2012].

That poses a problem to existing search algorithms: Such compounds could be of high interest for a search request, but how can be examined whether a compound comprises a given lexeme? A string match can be considered as an indication, but does not prove semantic relation. For, e.g., the stem *seh* [see] a simple string match would lead to an extensive list of words as can be seen in Figure 10, but by far not all of them show a semantic relation to sight. The same problem is faced when using lexicon based approaches where signal words are defined as lexemes only and need to be identified in all forms of appearance, and hence also as component of a compound.

FIGURE 10 SOME GERMAN WORDS THAT INCLUDE THE LETTERS SEH [SEE]

Especially for compounding languages, such as German, Dutch, Danish, Norwegian, Swedish, Russian or Finnish, compound handling is still a challenge – and even more when a word should be analyzed by its semantic components. This section explores the characteristics of compounds and their constituent elements for German, and compares seven algorithms with regard to runtime and error rates. More precisely, the subject of this section is the development, implementation and comparison of different algorithms that are able to decide whether a given German lexeme that occurs within a word is a complete semantic component of this word or just a part of another semantic component haphazardly containing the same letters. Searching e.g. for the semantic component *blau* [blue], one would want to have words like *kobaltblau* [cobalt blue] or *Blauhelm* [blue helmet] with a semantic relation to the word *blau* evaluated positively, whereas words like *Ablaufsteuerung* [sequence control system] or *halblaut* [in a low voice] with no relation to the word *blau* should be evaluated negatively.

The scope of this work is therefore different from typical stemming or decompounding approaches where the word components are unknown and need to be identified first. Even though one can use the same methods, the discussed problem is more concrete and can thus be solved faster and more efficiently with adequate tools. The aim of this section is to develop such algorithms that are easy to implement and to use, and to test them for their efficiency.

Such an algorithm can, e.g., be used to specify an additional filtering level of search results by identifying those words that do not only include the given lexeme as a string, but are also semantically linked to it. The result is a new subset *R* of relevant results within the search results *S* from a collection of words *W*. The relations between those sets are depicted in Figure 11. Furthermore, such an approach could be very useful for lexicon based methods categorizing text with the help of signal words that can occur in various appearance forms, such as affective [Köper and Schulte Im Walde] or perceptual [Kellner and Berendt 2011] vocabulary or idiomatic expressions [Citron et al. 2016].

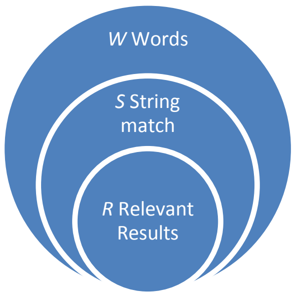

FIGURE 11 SUBSETS OF SEARCH RESULTS

To specify this set of relevant results *R*, we developed different algorithms based on word formation rules, some of them also using a dictionary and affix lists. The algorithms cover different approaches, ranging from phonetic to elaborate linguistic methods. For testing those algorithms, we chose (out of a list of several hundred lexemes) a set of 16 lexemes that induce different treatment in our algorithms and evaluated 1657 words containing one of these lexemes. We introduce these seven algorithms and compare their results, with regard to runtime and error rate.

4.4.1 Background and Related Research

Related research comes from the fields of morphology, phonology, and computational linguistics as well as information retrieval. First, we investigate the components of word formation and their structure in compounding languages, using German as example language. Second, we focus on stemming and compound splitting and take a look at methods and strategies used in this field.

4.4.1.1 Components of Word Formation

The structure of German words can be broken down to the following constituents:

- Lexemes are abstract word forms that express semantic content without giving grammatical information. We use the expression 'lexeme' as described by [Zemb 1984] as archilexeme, which includes all distinctive shared attributes of two semantically identic, but grammatically differently realized words; e.g. the allomorphic stems from the word *laufen* [to run] in its realizations *lauf*, *läuf* [run], and *lief* [ran].
- Affixes are morphemes that are attached to a word or its stem in order to form a new word. The most common forms are prefixes that are attached before, and suffixes that are attached after a word or its stem. Affixes can express grammatical information, but might also change the semantic content of a word.
- Flexion endings are put at the end of a word and express grammatical information. They can be split up into two groups. Declination endings are used for nouns and adjectives whereas conjugation endings are used for verbs.
- Word joints are linking morphemes that might occur within compounds and exist in paradigmatic (e.g. genitive or plural inflection) or nonparadigmatic form.

The German language is productive, which means that new words can be built by following existing morphemic patterns. The most important types of word formation are derivation, conversion and compounding. Derivation describes word formation with the help of affixes, especially prefixes and suffixes, and the use of Ablaut/Umlaut as markers for implicit derivation. Conversion or syntactical transposition describes the change of part of speech without any other morphological markers. Compounding is a process that combines at least two semantic components within one word. The place where different semantic components meet is called inner word border. It can be accompanied by a word joint.

All those morphemic patterns can be described by word formation models as the combination of one or more lexemes, affixes, word joints, and flexion endings. But description by morphemic components is not the only way to describe the structure of a word. Phonotactic models describe words as suites of consonant and/or vowel clusters and use statistical methods and rules in order to identify possible consonant/vowel clusters for a given language. The only word formation type that cannot be described that way is blending, which is the formation of words by building abbreviations; therefore we will not take that phenomenon into account.

4.4.1.2 Stemming and Compound Splitting

Stemming is a procedure that conflates word forms to a common stem. For example, the word forms *runner*, *runs*, and *running* are all reduced to *run* by a stemmer. Stemmers usually interpret white spaces and punctuation as word borders. This is a powerful method for non-agglutinative languages like English, where the components of ad hoc compounds tend to stay visibly separated by white spaces or hyphens, e.g. *compound verb idioms* or *long-lasting*, whereas one-word compounds are usually lexicalized. However, stemming does only cut away flection endings and some suffixes, but does not consider the inner structure of a word, which is problematic for long compounds such as the German word *Donaudampfschiffahrtsgesellschafts-kapitäne* [captains of Danube steamboat electrical services] that a stemmer could only reduce to its singular *…-kapitän*. At this point, compound splitting comes into use. Compound splitting analyzes a compound for its internal structure and splits a compound into its components [Escartín 2014].

Different strategies are pursued in order perform stemming, only some of them analyze the structure of compounds. Whereas brute force algorithms work with complete lookup tables that contain relations between root forms and inflected forms for all word forms (which is quite unrealistic for compounding languages such as German), suffix stripping algorithms make use of lists of suffixes that can be cut away in order to perform stemming [Porter 1980; Porter et al. 2002]. Morphologic algorithms work with morphologic description of words [Macherey et al. 2011]. Lemmatization algorithms use a more detailed grammar description as they first determine the part of speech of a word and then apply different normalization rules for each part of speech [Ingason et al. 2008]. Stochastic stemming algorithms are based on frequency and probability: with the help of unsupervised learning, huge corpora can be used as a reference in order to find similar solutions for a given word that needs to be stemmed or split up into its components [Patel and Shah 2016]. In their overview paper on decompounding in German, Braschler and Ripplinger [2004] analyze some more approaches like n-gram-based information retrieval and the use of well-engineered and mature stemmers like the NIST stemmer or the spider stemmer. Santos [2014] suggests to combine compound lists with other compounding methods due to runtime reductions.

One of the biggest challenges of natural language processing is ambiguity [Winkler 2015], especially on a morphological level [Bozsahin 2002]. Both stemming and decompounding should be executed in a well-balanced extent. In the case of 'overstemming', too much is cut away by a stemmer; if too little is cut away, one can speak of 'understemming'. For compound splitting, those effects are called 'aggressive' respectively 'conservative' decompounding [Singh and Gupta 2017].

Stemming and decompounding are used for processing user queries [Balakrishnan, Humaidi and Lloyd-Yemoh 2016] as well as in the field of machine translation [Clouet et al. 2015].

4.4.2 General Structure of the Algorithms

Even though our algorithms make use of different strategies, all algorithms have the same structure. Each algorithm is a realization of the parameterization as modeled in Figure 12. The input elements are a given string (a lexeme) and a word that includes that string. The verification process analyzes the semantic relation between the word and the given string. The output is Boolean (true/false) and consists of a decision about that semantic relation of the two input parameters.

We developed seven different algorithms. Besides the naïve search algorithm, all are based on word formation rules, some of them also using a dictionary and affix lists. They all have the same structure and use the same interface.

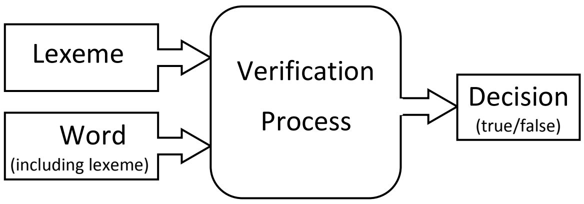

FIGURE 12 PARAMETERIZATION OF THE ALGORITHMS

As an input, all algorithms need a given lexeme and a word that includes that lexeme. Some algorithms need additional information concerning the possible part of speech (further addressed as PoS) realizations of the lexeme. Those options are coded binary, considering that one lexeme can have several possible PoS realizations (e.g., *ärger* [trouble] can be realized as an adjective, a noun or a verb). We only took the three most current PoS realizations of the German language into account, namely adjectives, nouns and verbs. Table 7 shows the different PoS realization types and their codification.

C	Possible PoS realization
1	adjective
2	noun
3	adjective, noun
4	verb
5	adjective, verb
6	noun, verb
7	adjective, noun, verb

TABLE 7 PART OF SPEECH (POS) REALIZATION TYPES

4.4.2.1 Problem Description

The starting point for each algorithm at the moment of its function call is illustrated in Figure 13. It shows two compounds containing the string *röt* that can be produced by vowel gradation on the lexeme *rot* [red].

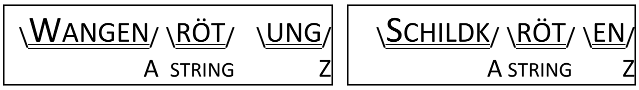

FIGURE 13 DECISION PROBLEM

As the input arguments are a lexeme and a word that contains that lexeme, we know the start position of the lexeme in the word and what is before (= part “A”) and after (= part “Z”) that string. The mission of each algorithm is now to decide whether A and Z are valid letter combinations, built on the respective word grammar model. Looking at our examples, *Wangenrötung* [cheek reddening] should be evaluated positively, whereas *Schildkröten* [turtles] should be evaluated negatively.

4.4.2.2 Resources

The algorithms can use resources that are stored in a database. It contains

- a dictionary of all German word forms: Each lexical entry is listed several times in all possible declination and conjugation versions.

- a list of the most frequently used affixes of the German language. Each entry contains the following information:
  - the string of the affix
  - the PoS it can be used with (see the codes in Table 1)
  - the PoS it produces (see the codes in Table 1)
  - the affix type (word joint = 1, prefix = 2, suffix = 4)
  - and a specification on the suffix type (not a suffix = 0, suffix = 1, ending = 2, suffix or ending = 3).

  The entry for the suffix *lich*, e.g., is stored as *lich, 7, 1, 4, 1*.

4.4.3 The Algorithms

In this section, we introduce our algorithms. The algorithms cover different approaches, ranging from phonetic to elaborate linguistic methods. While algorithm 0 only performs naïve string search, algorithm 1 and 2 are realizations of different morphologic word description models. Algorithm 3 is a lemmatization algorithm. Algorithm 4 uses suffix stripping; algorithm 5 is based on phonotactic rules. Algorithm 6 is a realization of a stochastic stemming algorithm that uses unsupervised learning. If the respective method does not cover the analysis of important word components, the method is completed by operational sequences defined by previous algorithms; e.g. suffix stripping only covers suffix handling and has no rules for prefix handling, which are taken from algorithm 1.

4.4.3.1 Algorithm 0

Algorithm 0 is a naïve search algorithm that tests only whether a word contains the letters of the given lexeme or not. As soon as that is the case, the request is evaluated positively. It is used as a baseline.

4.4.3.2 Algorithm 1

Algorithm 1 uses an affix list and a dictionary and tests if the letters before and after the given lexeme might be split up into a composition of affixes and dictionary entries. If this is the case, the word is evaluated positively.

4.4.3.3 Algorithm 2

Algorithm 2 uses a better word formation model than algorithm 1 and is based on [Kellner 2010].

Kellner differentiates two sorts of suffixes: “preendings” that occur directly after a root word and “endings” that are a result of conjugation or declination and indicate the end of a word. Any word can hence be described as a regular expression[9] with the following structure:

((prefix) (root word){1} (preending)* (word joint)?)?*
(prefix) (root word){1} (preending)* (ending)?*

The structure of our example *Wangenrötung* can thus be analyzed as *(root word)(word joint)(root word)(preending)*.

For comparison reasons, our realization is a simplified version and does not handle exceptions. The rest works in the same way as algorithm 1.

4.4.3.4 Algorithm 3

Algorithm 3 uses a system similar to algorithm 2, but is extended by a rule set deciding which PoS realization type may be combined with which affixes. E.g., the suffix *bar* [-able] can only be combined with PoS realization type 4 (verbs) and produces an adjective. In contrast to previous algorithms, this one uses all the information stored in the affix lists (structure to be found in Section 4.4.2) and needs, as additional input, the PoS realization type of the given string (to be found in Table 7).

4.4.3.5 Algorithm 4

Algorithm 4 is based on Porter’s stemming algorithm [Porter 1980], respectively on its adaptation for German [Porter et al. 2002] which describes a word as a sequence of consonant and vowel clusters. Porter’s stemming algorithm does without complex description of word grammar. It simply divides words into vowel (V) and consonant (C) clusters which are defined as a suite of at least one letter of the same class (either vowels or consonants). Each word can hence be reduced to a regular expression of the following structure:

(C)? (VC){m} (V)?

[9] Brackets define the application area of a quantifier. Quantifiers define how often an element may be repeated: * = unlimited times, ? = 0 or 1 times. Other numbers of repetition are numbers within curly braces, e.g. {1} = exactly once, {3-5} = 3, 4 or 5 times, etc.

The word *Schildkröte* [turtle] can thus be notated as *CVCVCV (m=2)* and *Pfau* [peacock] as *CV (m=0)*. From that structure, the algorithm cuts away predefined suffixes under given circumstances. The result is a (somehow) stemmed word without conjugation and declination endings. Porter's algorithm is not very aggressive due to overstemming problems in the context of information retrieval and it only describes suffix stripping.

For comparison reasons, we combine Porter's method with our previous strategies by first executing Porter's suffix stripping and then looking up the remaining letters in our suffix lists and the dictionary.

4.4.3.6 Algorithm 5

Algorithm 5 is based on phonotactic rules and is the only one not using a dictionary.

Phonotactics is a subfield of phonetics that describes the permissible combinations of phonemes for a given language [Carstensen et al. 2009]. Consonantal phonemes can be combined to consonant clusters considering phonotactic rules. In German, e.g., the phoneme *g* can be followed by an *l* like in *glauben* [believe], but not by a *p*. Still, such a consonant combination can occur at the inner word borders of a compound, like in *Bergpass* [mountain pass] or *Fertigprodukt* [convenience product].

The algorithm works analogously to the first algorithms concerning affixes, but instead of the dictionary lookup it tests whether the last consonants before and the first consonants after the given string are in the list of known consonant clusters for the German language. If both parts are found, the word is evaluated positively.

4.4.3.7 Algorithm 6

Algorithm 6 uses the method of supervised learning. All words are described as a composition of up to three word components before and up to three word components after the signal word. The algorithm needs to be trained on a set of words where those components are explicitly defined and works with a decision tree that looks for the word in the training set that corresponds in structure best to the word in processing.

In order to find a common structure for all words, all words from the training corpus need to be annotated according to the schema that can be found in Figure 14.

A1 A2 A3 L Z1 Z2 Z3

FIGURE 14 WORD FRAGMENTATION

Each word fragment is classified and assigned a value according to Table 8.

Fragment	Classification
L (Given Lexeme)	1 = adjective
	2 = noun
	4 = verb
A1-3 and Z 1-3	0 = empty
	1 = word joint
	2 = prefix
	3 = dictionary
	4 = suffix
	5 = none of

TABLE 8 WORD FRAGMENT CLASSIFICATION

Shorter words are surrounded by word fragments with the value 0. The value 5 is assigned to word fragments that cannot be identified as an affix or a dictionary entry and leads to a negative evaluation.

4.4.4 Implementation

In this section, we describe how we implemented the algorithms presented previously, looking at structure, input and output, and the database and its components.

As already described on an abstract level in section 3, all algorithms are built with the same structure. As an input, all algorithms need a given lexeme and a word that includes that lexeme. Valid calls would, e.g., be `eval(Kind, Kindergarten)` [child, kindergarden] or `eval(rot, Roggenbrot)` [red, rye bread]. The correct output would be `true` for the first and `false` for the second call.

Valid calls for the algorithms that need additional information concerning the possible PoS realizations of the given string need one more argument and are therefore, e.g., `eval(Kind, Kindergarten, 2)` [child, kindergarden] or `eval(rot, Roggenbrot, 1)` [red, rye bread].

Once called, the program communicates with our database via interface. Our database contains not only a dictionary of all German word forms and the affix list, but also a list of search argument strings and the test corpus where the search is performed on.

Our affix lists for the German language were built upon the affix lists by [Kellner 2010] and [Canoo Engineering AG 2011] and expanded by us. They now contain 116 affixes.

The list of permissible consonant clusters for the German language was built upon [Jürgenson 2009] and [Hess 2002] and includes 20 consonantal phonemes, 30 consonant clusters for the beginning and 72 for the end of a word.
We used a German dictionary containing about 240,000 entries. We didn't take capitalization into account.

Figure 15 shows a class diagram with the operations that can be used for interaction with the database.

INTERFACE DB
`dictionaryContainsWord (String word)`: bool `containsAffix (String word)`: bool `getRule (String affix, int type)`: int `getAffix (int type)`: ResultSet `getAffixClass (int type, int class)`: ResultSet

FIGURE 15 CLASS DIAGRAM: DATABASE INTERFACE

Interaction with the database is possible using the following commands:

- `dictionaryContainsWord()` looks up a given string in the dictionary.
- `containsAffix()` looks up a given string in the affix lists.
- `getRule()` returns which PoS class may be created with a given affix and a PoS realization type
- `getAffix()` returns a list of possible affixes for a given PoS realization type
- `getAffixClass()` returns a list of possible affixes for a given PoS realization type and a PoS class

All algorithms were implemented in PHP 5.2.4. We used MySQL 5.0.51a as a database. All scripts were executed on an Apache 2 webserver on Ubuntu 8.04, 2.2GHz, 2GB RAM.

4.4.5 Evaluation

In this section, we describe the complete evaluation process. First, we explain what elements were chosen for evaluation purposes and how the test corpus was set up. Second, we define our measures. Third, we present the evaluation results, which are then discussed. We also discuss limitations of the results and our approach.

4.4.5.1 Test Corpus

For the evaluation, we need two corpora: on the one hand, we need a list of given lexemes that show the differences between our algorithms; on the other hand, we need a text corpus where those lexemes can be found in all their appearance forms.

The list of lexemes was built choosing lexemes based on following conditions:

- the lexeme should be a single unit of meaning
- the same string should also appear in words with a different meaning
- evaluation results for the given lexeme should not be the same for all seven algorithms.

Like Geyken and Hanneforth [2006], we have decided to handle irregular morphology like Ablaut in the lexicon. For example, the irregular verb *riechen* [smell] is stored in its two allomorphic states *riech* (present stem) and *roch* (past stem). We started with a handcrafted list with several hundred lexemes, which also appear in words with a different meaning. After the first round of evaluation, we excluded all lexemes that were evaluated with the same results by all our algorithms (either all true or all false) and only kept lexemes that caused inhomogeneous evaluation results. That led to a number of 16 lexemes, which can be found in Table 9.

We used that list on a text corpus in order to identify different words including the lexeme or its string. From a German corpus with 3 million phrases[10], we filtered out all words that contain one of the 16 strings, which was the case for 1800 different words[11]. In a next step we manually excluded proper names and foreign words, which led to a list of 1657 remaining words.

[10] http://corpora.uni-leipzig.de/downloads/deu_news_2008_3M-mysql.tar
[11] ‘Different’ is meant as a difference of at least one letter.

Of course, evaluation on such a corpus necessarily leads to a very high error rate. This is chosen with intent to show the bigger differences between the algorithms that are subject of interest. Both corpus and list of lexemes were stored in the database.

Semantic unit	Classification	English Translation
biss	N/V	bit/bite
blau	A	blue
bläu	A	blue
find	V	find
flach	A	plane
hell	A	lucid
hör	V	hear
kling	V	sound
lampe	N	lamp
laut	N	loud
riech	V	smell
roch	V	smell
rot	A	red
röt	A	red
sah	V	see
seh	V	see

TABLE 9 LEXEMES CHOSEN FOR EVALUATION

Furthermore, a training set was needed for the algorithm using supervised learning. Algorithm 6 was trained on a training set of 642 entries out of the 1657 that were annotated manually by the authors, according to the structure as described in Section 4.6. The training set was defined manually and set up in two steps: First, the semantic relation between each word and the given lexeme was evaluated manually, e.g. *[Wangenrötung, röt, 1]* or *[Schildkröten, röt, 0]*. Second, each word was annotated manually according to the word fragmentation schema (as described in Figure 14). The decision tree is induced according to [Gupta 2006].

4.4.5.2 Measures

Each algorithm is executed separately. The evaluation results from each algorithm are stored together with their runtime. Algorithms can hence be compared with regard to runtime and error rates in comparison to the manually defined results.

Errors can be divided into two different groups:

- Error #1: The evaluation result by the algorithm is false positive, which means evaluation is positive whereas the correct evaluation result is negative.

- Error #2: The evaluation result by the algorithm is false negative, which means evaluation is negative whereas the correct evaluation result is positive.

For query analysis, error #1 is the somehow less serious problem; in worst case, some not relevant search results are shown to the user who has to identify relevant hits anyways. Error #2 is the more problematic error as relevant search results are sorted out and cannot be accessed by the user.

The error rate is calculated as the sum of errors of both classes.

Runtime was calculated as the arithmetic mean of ten runs for each algorithm. The measures for runtime are relative values: Algorithm 0 is the fastest; its runtime is normalized to 1. The runtime for the other algorithms is expressed in relative values, e.g. algorithm 5 is 20 times slower than algorithm 0.

4.4.5.3 Results

This section shows how well each algorithm works and discusses the findings.

Each algorithm is analyzed for its error rate and for runtime. The error rate (allover, error #1, error #2) is given in absolute and relative numbers. The evaluation results are shown in Table 10.

	Error rate		**Error #1**		**Error #2**		**Runtime**
	abs.	%	abs.	%	abs.	%	
Algorithm 0	798	48,16	798	48,16	0	0	1
Algorithm 1	285	17,2	156	9,42	129	7,79	79,02
Algorithm 2	271	16,35	193	11,65	78	4,71	60,92
Algorithm 3	251	15,15	147	8,87	104	6,28	122,27
Algorithm 4	263	15,87	165	9,96	98	5,91	209,02
Algorithm 5	489	29,51	474	28,61	15	0,91	20,11
Algorithm 6	289	17,44	136	8,21	153	9,23	66,28
Number of words in the evaluation: 1657							

TABLE 10 ALGORITHM EVALUATION RESULTS

Out of 1657 words in our corpus, by manual evaluation, 859 were thought to be evaluated with a positive and 798 with a negative result.

908 words were evaluated with the correct result by all algorithms (except the naive algorithm 0), of which 602 were correctly evaluated positively and 306 negatively.

Correct positive evaluation was mostly accomplished for simple compounds with a maximum of two constituents (with or without word joint) or one suffix. Very few words with more than one suffix were evaluated correctly by all algorithms, which is not surprising as most of the algorithms use the same procedure for prefix handling. The number of prefixes does not seem to influence the result. Correct negative evaluation was mostly accomplished for strings that occur haphazardly in a word, which is the case for e.g. *rot* [red] in *Brot* [bread], *Protokoll* [protocol] or *Schrott* [scrap], or for *riechen* [smell] in *kriechen* [crawl] or *Griechen* [Greek]. Also most of the compounds including such strings were evaluated correctly.

However, there are several words where all algorithms failed and produced error #1 results, e.g. *rot* [red] in *neurotisch* [neurotic]; *seh* [see] in *Sehne* [tendon], *sehnen* [yearn] or compounds with the word joint *s* plus *Ehe* [marriage]; or *kling* [sound] in *Klinge* [blade]. The problem lies in the fact that for those words, the letters around the given lexeme can be matched completely with affix and dictionary entries and are hence evaluated positively.
No words evoked error #2 results in every algorithm.

4.4.5.4 Discussion

Due to our decision to only include difficult words in our evaluation, the error rates are much higher than they would be when analyzing standard corpora.

Comparison with the naïve algorithm 0 shows that error rates can be drastically decreased with the help of intelligent evaluation algorithms. Already the phonetic evaluation (algorithm 5) can reduce the error rate for about 40 percent of algorithm 0's error rate whereas the dictionary based algorithms (algorithms 1-4, 6) show error reductions to only a third.

A detailed analysis of the results showed the following strengths and weaknesses of specific algorithms:

Algorithm 1 and 2 do not show any special dissenting results. This may easily be explained by the fact that most of the other algorithms are built upon the structure of those two algorithms.

Algorithm 3 is performing best concerning the absolute error rate because of its rules for suffix use. At the same time, those rules are not covering all cases and may sometimes lead to an exclusion of relevant matches (error #1), e.g. the suffix *ung* [-ing] may, according to our classification concerning the PoS realization type, only be combined with verbs and nouns and hence excludes *Rötung* [reddening] or the participle *erhellt* [illuminated], which actually would demand a more complex definition of state changes because *erhellt* cannot be deduced directly to *hell* [lucid], but only to the verb *erhellen* [illuminate].

Algorithm 4 was the only algorithm that produced an error #2 for the verb *klingeln* [to ring] and all its derivatives in relation to the semantic unit *klingen* [sound].

Algorithm 5 has a really low number of error #2 results, but a very high number of error #1 results because of the positive evaluation of lexemes that are surrounded by valid consonant clusters, e.g. for *rot* [red] *Karotte* [carrot], *Erotik* [erotic], or *Knäckebrot* [crispbread]. Especially the last example shows the problematic reduction to consonant clusters: *Knäckebrot* (consonant cluster: *b*) in relation to *rot* is evaluated incorrectly positively whereas *Roggenbrot* [rye bread] (consonant cluster: *nb*) oder *Brot* [bread] (consonant cluster: *b*, but without any preceding vowel) are evaluated negatively.

Algorithm 6 shows some errors due to an apparent lack of adequate examples in the training set. The problem yet doesn't lie in the algorithm, but in the composition of the training set.

The results for runtime are thought as a rough orientation and should not be misinterpreted as disqualification criterion for the slower algorithms. On the one hand, the absolute runtime lies between 0.05 and 10.50 seconds, which remains within reasonable bounds for all algorithms. On the other hand, the performance strongly depends on the implementation of the algorithms, the used programming language and the technical equipment; e.g. algorithm 3 might perform better in a logic oriented programming language, and algorithm 4 would be likely to perform better if implemented in a programming language with better string manipulation than PHP.

Overall, the choice of the algorithm should depend on the specific problem that needs to be solved. If it is crucial to include all potentially relevant search results, one should prefer an algorithm with a low error #2 rate, which is the case for algorithm 5. In cases where the number of search stems is quite small, one could use one of the algorithms described above and assign exception lists manually. However, for larger stem sets a combination of several approaches seems to be promising.

4.4.5.5 Limitations

Of course, our grammatical approach can point out semantic relation only to a certain extent. In our approach, we do not take semantic changes caused by the combination of the lexeme with different affixes or lexemes into account. The word *sprechen* [speak] can, e.g., occur as *vorsprechen* [audition] or *nachsprechen* [repeat sth. after sb.] where the semantic relation is perfectly clear. But this is not any more the case for *entsprechen* [comply], which still could be reduced to *sprechen* when only looking at the grammatical aspect, but actually has little relation to the verb *sprechen* in means of a communicative act.

Another reason for misclassification can be rule-consistent, but incorrect compound splitting according to grammatical schemata; e.g. *neurotisch* [neurotic] has no semantic relation to *red* [rot], but can be incorrectly split up into the semantic units *neu* [new] + *rot* [red] + *isch* [ic].

Such problems could partly be solved by definition of exceptions and exception rules as suggested by [Kellner 2010]. Such an approach demands an extensive investigation of manual work, but would make the error rate decrease, especially for errors #1.

Another topic that has not been solved with our algorithms is the handling of polysemy: compounds that are ambiguous either because they include ambiguous components or because their semantic content depends on their pronunciation, would need some special treatment. As our algorithms only work on a lexical basis, we did not take disambiguation problems into account. This could of course be tackled when moving from a lexical to a syntactical examination level.

4.5 The Signal Term Extraction Method (STEM) Algorithm

After testing several algorithmic approaches, we describe our STEM algorithm. STEM stands for “Signal Term Extraction Method”. It deals with the problems described above and tests for every word within a text whether it is a valid word realization (any part of speech, simplex, compound, conjugated or declined) of a stem from a given signal stem list or just another term haphazardly containing the same letters, and counts the occurrences of signal stems categorized by predefined semantic dimensions. It is based on algorithm 2 [Kellner 2010; Kellner and Grünauer 2012] and extended by several additional components.

The STEM algorithm is built upon 4 main algorithmic blocks (lookup, stem identification, surrounding letters check, counter) and 4 external data resources (signal stem list, tagged terms database, word component lists, dictionary of all word forms). In short, the algorithm works as follows: The input text is analyzed term by term; each term is first looked up in the tagged terms database, which contains all previously tagged word realizations and their semantic dimension. If the term is found there, the program can jump directly to the counter and, after adding 1 to the relevant semantic dimension, go on with the next term. If the term is not found in the tagged terms database, it is sent to the stem identification where the word is shortened iteratively, until either a stem from the VAKOG stemming list is identified or only 2 letters are left. If a match is found, the remaining letters are sent to the surrounding letters check. If all of the remaining letters can be covered with elements from the word component lists and the dictionary of all word forms, the term is sent to the counter, otherwise it is sent back to the stem identification, which, if no other positive match is found, sends the term back to the counter as well. A diagram of the STEM algorithm and its components can be found in Figure 16.

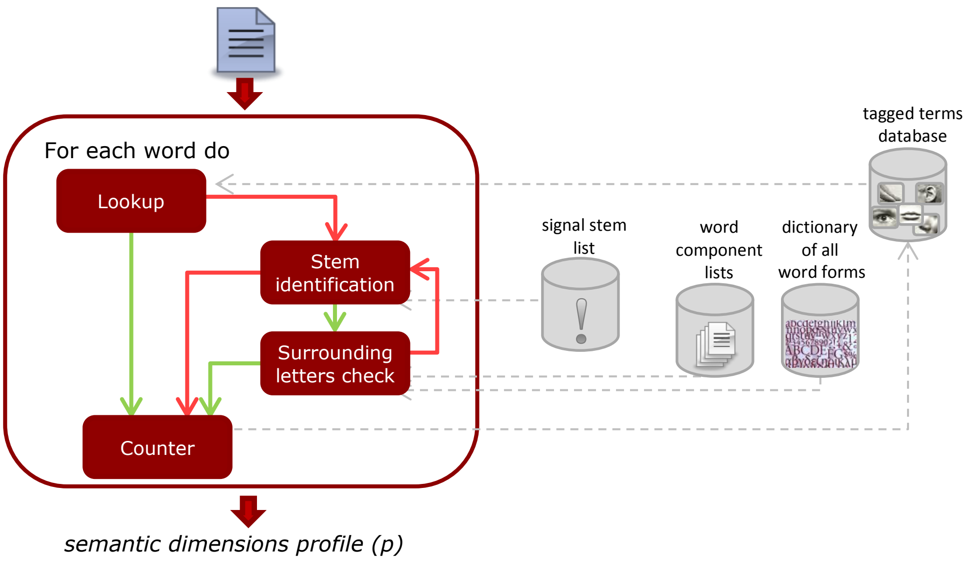

FIGURE 16 THE STEM ALGORITHM AND ITS COMPONENTS

4.5.1 Operations and Components

The STEM algorithm components shall now be explained in detail.

Input can be any text. It is cut into junks of single terms, by use of white spaces and punctuation marks as separators. Then, each term runs through the following steps.

In the **Lookup**, the term is looked up in the ***tagged terms database***, which contains all previously tagged word realizations and their semantic dimension. In our case, those semantic dimensions are the four perceptual categories of visual (V), auditory (A), kinesthetic (K) and olfactory/gustatory (OG) vocabulary, furthermore a non-VAKOG (X) dimension that contains all words that do not belong to any of the semantic dimensions above (e.g., *sichtbar* [visible], *V* or *nichts* [nothing], *X*). If the term is not found there, the term is sent to the stem identification unit. As a result, only words not yet included in the tagged terms database need further analysis. Every time a new word is classified, the tagged terms database is expanded automatically. As a result, the process of scanning a text with regard to the occurrence of perceptual vocabulary is much faster than if every word needs to run through the whole process of classification. If the term is already found in the tagged terms database, the program can jump directly to the counter.

The **Counter** sums up the number of terms per semantic dimension as defined in the tagged terms database. As soon as the semantic dimension of the term is identified (either by lookup or after the stem identification), the counter adds 1 to the relevant semantic dimension and goes on with the next term. The perceptual indicator of the term is calculated as follows:

indicator level p(t) = 1 if stem(t) is in the lexicon of perceptual vocabulary
= 0 in all other cases.

During the **Stem identification**, the word is shortened iteratively from both sides until either a stem from the ***signal stem list*** is identified or only 2 letters are left. The signal stem list contains all tagged stem realizations and their semantic dimension (e.g., *seh*, *säh*, *sah*, *sieh*, *sicht* for the verb *sehen* [see], each one stored with the attribute *V*), using the same dimensions as described for the tagged terms database. In the case of no match, the term is sent back to the counter with an indicator level 0, otherwise an array with all possible matches is sent to the surrounding letters check and being tested, beginning with the longest match.

The underlying model of word grammar can be described as follows: A term is handled as a valid word realization if all surrounding letters can be covered according to the schema presented in Section 4.4.3 for algorithm 2.

((prefix) (root word){1} (preending)* (word joint)?)?*
(prefix) (root word){1} (preending)* (ending)?*

where one root word is covered by the match from the signal stem list, and the remaining letters shall be covered with elements from the word component lists and the dictionary of all word forms.

The word component lists are four lists, one per word component, namely prefixes, preendings, endings, and word joints. These word components are based on a prior publication by the author [Kellner 2010] and have been extended since.

A prefix can be defined as an affix placed before a base or another prefix, as *un-* in *unkind*, *un-* and *re-* in *unrewarding*.

a	*darein*	*fehl*	*hinab*	*ko*	*prä*	*voran*
ab	*davon*	*fort*	*hinan*	*kol*	*pro*	*voraus*
abwärts	*davor*	*ge*	*hinauf*	*kom*	*pseudo*	*vorbei*
an	*dazu*	*geg*	*hinaus*	*kon*	*re*	*vorher*
anti	*dazwischen*	*haupt*	*hindurch*	*kor*	*Rück*	*weg*
auf	*de*	*her*	*hinein*	*los*	*sub*	*wider*
aufwärts	*des*	*herab*	*hinter*	*mega*	*super*	*wieder*
aus	*dis*	*heran*	*hinunter*	*miss*	*tief*	*zer*
auseinander	*drein*	*herauf*	*hinweg*	*miß*	*top*	*zu*
be	*durch*	*heraus*	*hinzu*	*mit*	*trans*	*zurecht*
bei	*ein*	*herbei*	*hoch*	*nach*	*über*	*zurück*
bevor	*einher*	*herein*	*hyper*	*nahe*	*ultra*	*zusammen*
da	*empor*	*herüber*	*il*	*näher*	*um*	*zwischen*
dabei	*ent*	*herum*	*im*	*neben*	*umher*	
daher	*entgegen*	*herunter*	*in*	*nicht*	*un*	
dahin	*er*	*hervor*	*innen*	*nieder*	*unter*	
dar	*erz*	*herzu*	*inner*	*non*	*ur*	
daran	*ex*	*hierher*	*inter*	*para*	*ver*	
darauf	*extra*	*hin*	*ir*	*post*	*vor*	

FIGURE 17 WORD COMPONENTS: PREFIXES

A preending is an affix that follows the element to which it is added, as *-ly* in *kindly*. It is an artificial class that combines confixes, suffixes and suffixoids as well as participle and comparison forms (comparative and superlative).

abel	*atur*	*er*	*i*	*ion*	*los*	*schaft*
ade	*bar*	*erei*	*ian*	*isch*	*mals*	*st*
age	*bold*	*erie*	*iat*	*isier*	*mann*	*t*
al	*chen*	*erin*	*iatin*	*ismus*	*maßen*	*tel*
and	*de*	*erisch*	*ibel*	*ist*	*mäßig*	*tum*
ant	*dings*	*erl*	*ice*	*istin*	*ment*	*um*
antin	*e*	*ern*	*ie*	*istisch*	*n*	*ung*
anz	*ei*	*esk*	*iell*	*it*	*ner*	*ur*
ar	*el*	*eur*	*ier*	*ität*	*nerin*	*us*
är	*elchen*	*esse*	*ifizier*	*iv*	*nis*	*wärts*
arin	*ell*	*ette*	*ig*	*izität*	*oid*	*weg*
ärin	*ement*	*euse*	*igkeit*	*jan*	*or*	*werk*
ast	*en*	*fach*	*ik*	*keit*	*orin*	*weise*
at	*end*	*fähig*	*iker*	*lei*	*os*	*wesen*
atin	*ens*	*frau*	*ikerin*	*lein*	*ös*	*würdig*
ation	*enswert*	*haft*	*in*	*ler*	*rich*	*zeug*
ativ	*ent*	*halben*	*ine*	*lerin*	*s*	
ator	*entin*	*halber*	*ing*	*lich*	*sal*	
atorin	*enz*	*heit*	*innen*	*ling*	*sam*	

FIGURE 18 WORD COMPONENTS: PREENDINGS

An ending can be a verb conjugation or a noun or an adjective declination ending.

e	*er*	*et*	*n*	*st*	*tem*	*tes*
em	*ern*	*in*	*nd*	*t*	*ten*	*test*
en	*es*	*innen*	*s*	*te*	*ter*	*tet*

FIGURE 19 WORD COMPONENTS: ENDINGS

A word joint is a unit, which can occur as phonological unit between the parts of a compound (e.g. *Volk-s-musik* [folk music]).

e	*ens*	*es*	*ns*
en	*er*	*n*	*s*

FIGURE 20 WORD COMPONENTS: WORD JOINTS

The Surrounding letters check tries to cover all remaining letters around the match from the signal stem list with entries from the word component lists and the dictionary of all word forms. First, the letter sequence before the match from the signal stem list (part A) is looked up in the prefix list and, if there is no match, also in the dictionary of all word forms (taking word joints into account). If there is no match, part A is shortened letter by letter from the right-hand side, until the remaining part is either found in the prefix list or too short for further lookup. For the first case, the letters that could be matched with a prefix entry are removed. The remaining letters are sent to the beginning of this paragraph and looked up again in the prefix list and in the dictionary. Again, they might be shortened iteratively, but all remaining letters of part A need to be matched. The procedure for part Z (the letter sequence after the match from the signal stem list) is executed analogously, with a differentiation between preending and ending: First, part Z is looked up in the preendings and the endings list (both can be the last part of a word) and, if there is no match, also in the dictionary of all word forms. If there is no match, part Z is shortened letter by letter from the right-hand side, until the remaining part is either found in the preendings list (preendings can occur repeatedly) or is too short for further lookup. For the first case, the letters that could be matched with a preending entry are removed. The remaining letters are sent to the beginning of this paragraph and looked up again in the preendings and endings list and in the dictionary. Again, they might be shortened iteratively, but all remaining letters of part Z need to be matched. The endings list can only be used for matching the very last letters of Z because an ending finishes a word and can only occur once. The operational sequence is expressed as a (slightly simplified) flowchart diagram in Figure 21.

If all remaining letters are covered, the result is true and the term is sent to the counter with an indicator level 1 for the dimension assigned to the stem from the signal stem list that led to a full cover, else, the indicator level is 0. The so newly analyzed term is stored as a new entry in the tagged term database.

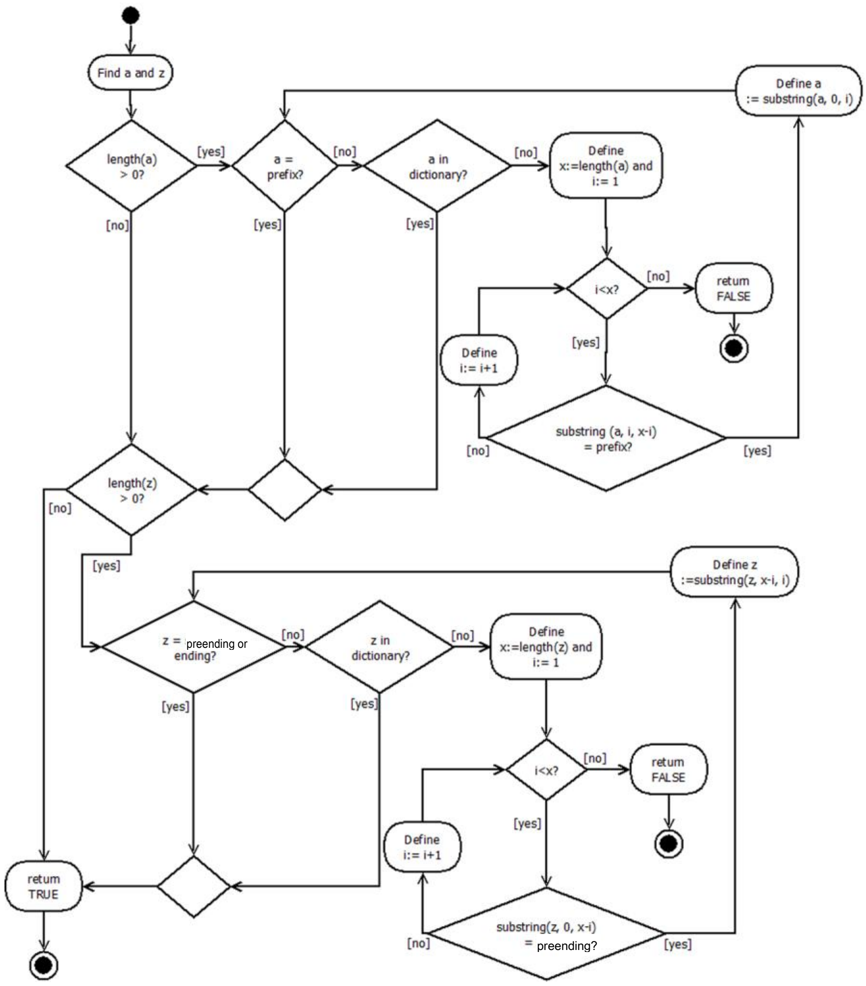

FIGURE 21 FLOWCHART DIAGRAM FOR THE SURROUNDING LETTERS CHECK

4.5.2 Execution Examples

The evaluation of the word *Wangenrötung* [cheek reddening] starts with a lookup in the tagged term database. If it is not there already, the word is sent to the stem identification where the word is shortened iteratively.

WANGENRÖTUNG	ANGENRÖTUNG	NGENRÖTUNG	GENRÖTUNG	ENRÖTUNG	NRÖTUNG	RÖTUNG	ÖTUNG	TUNG	UNG
WANGENRÖTUN	ANGENRÖTUN	NGENRÖTUN	GENRÖTUN	ENRÖTUN	NRÖTUN	RÖTUN	ÖTUN	TUN	UN
WANGENRÖTU	ANGENRÖTU	NGENRÖTU	GENRÖTU	ENRÖTU	NRÖTU	RÖTU	ÖTU	TU	
WANGENRÖT	ANGENRÖT	NGENRÖT	GENRÖT	ENRÖT	NRÖT	**RÖT**	ÖT		
WANGENRÖ	ANGENRÖ	NGENRÖ	GENRÖ	ENRÖ	NRÖ	RÖ			
WANGENR	ANGENR	NGENR	GENR	ENR	NR				
WANGEN	ANGEN	NGEN	GEN	EN					
WANGE	ANGE	NGE	GE						
WANG	ANG	NG							
WAN	AN								
WA									

FIGURE 22 EXECUTION EXAMPLE: WANGENRÖTUNG

Only one match is found in the signal stem list, namely *röt* [red], which is assigned to the perceptual dimension V. The term is now sent to the surrounding letters check. Part A, *Wangen* [cheeks], is searched in the prefix list, without a match; and then, successfully, in the dictionary of all word forms. Part Z, *ung* [-ing], is looked up in the preendings list and is found as a whole, no need to check the endings list or the dictionary. As both parts can be matched, the evaluation is positive. A new entry is sent to the tagged terms list (*Wangenrötung*, *V*), the counter for V is augmented by 1.

Also the word *Schildkröten* [turtles] is sent to the stem identification if it cannot be found in the tagged terms database. There, it is shortened iteratively.

SCHILDKRÖTEN	CHILDKRÖTEN	HILDKRÖTEN	ILDKRÖTEN	LDKRÖTEN	DKRÖTEN	KRÖTEN	RÖTEN	ÖTEN	TEN
SCHILDKRÖTE	CHILDKRÖTE	HILDKRÖTE	ILDKRÖTE	LDKRÖTE	DKRÖTE	KRÖTE	RÖTE	ÖTE	TE
SCHILDKRÖT	CHILDKRÖT	HILDKRÖT	ILDKRÖT	LDKRÖT	DKRÖT	KRÖT	**RÖT**	ÖT	
SCHILDKRÖ	CHILDKRÖ	HILDKRÖ	ILDKRÖ	LDKRÖ	DKRÖ	KRÖ	RÖ		
SCHILDKR	CHILDKR	HILDKR	ILDKR	LDKR	DKR	KR			
SCHILDK	CHILDK	HILDK	ILDK	LDK	DK				
SCHILD	CHILD	HILD	ILD	LD					
SCHIL	CHIL	HIL	IL						
SCHI	CHI	HI							
SCH	CH								
SC									

FIGURE 23 EXECUTION EXAMPLE: SCHILDKRÖTEN

Like for the first example, only one match is found in the signal stem list, namely *röt* [red], which is assigned to the perceptual dimension V. The term is now sent to the surrounding letters check. There, neither the prefix nor the dictionary lookup for part A, *Schildk*, is successful. The first iteration of letter shortening leads to a dictionary match with *Schild* [shield]. The remaining letter *k* is looked up in the prefix list, but cannot be found. As already part A cannot be reconstructed as a combination of dictionary entries (and neither be combined with a word joint) and/or prefixes, the evaluation leads to a negative result.

4.6 The VAKOG Analysis Pipeline[12]

The Signal Term Extraction Method (STEM) Algorithm needs, of course, to be embedded into a pipeline of other analyses steps. In order to choose the right ones, we will first take a look at the measures that lie behind the calculations and the hypotheses that shall be tested with the help of the pipeline.

4.6.1 Measures

Let T denote the text, and t an embedded linguistic unit, a term, where $t \in T$. Let the set of sensory classes be defined as $S = \{s_1, \ldots, s_5\} = \{V, A, K, OG, X\}$, where the first four stand for visual, auditory, kinesthetic, olfactory and gustatory dimensions, while X denotes the special case of *neutrality*, or absence of sensory reference. The goal is to determine a mapping function $f : m \rightarrow s_i$, such that we obtain an ordered labeled pair (m, s_i). The mapping is based on the tagged terms database, which is partly filled in advance and continuously expanded with the help of the Signal Term Extraction Method (STEM) Algorithm. For the very rare case that multiple sensory classes can characterize m, then the first one that is classified by the algorithm is chosen.

The use of perceptual vocabulary in a post p is expressed as a four-dimensional vector *vprofile(p) = [p.V, p.A, p.K, p.OG]* based on the frequency of perceptual vocabulary per perceptual system. Thus, the components of the vector are defined as

$$p.V = \frac{|\{ t \in p' \mid v(t) = 1\}|}{|\{t \in p' \mid v(t) = 1 \text{ or } a(t) = 1 \text{ or } k(t) = 1 \text{ or } og(t) = 1\}|}$$

$$p.A = \frac{|\{ t \in p' \mid a(t) = 1\}|}{|\{t \in p' \mid v(t) = 1 \text{ or } a(t) = 1 \text{ or } k(t) = 1 \text{ or } og(t) = 1\}|}$$

[12] This section is based on the publications „Towards a New Dimension for User Modeling: The Use of Sensory Vocabulary“ [Kellner and Berendt 2012], which was awarded with the „Best Poster Award“ at the UMAP 2011 (User Modeling, Adaptation and Personalization) in Girona, Spain; and on its extended version „Extracting Knowledge about Cognitive Style. The Use of Sensory Vocabulary in Forums: A Text Mining Approach“ [Kellner and Berendt 2011], which was awarded with the „Best Paper Award“ at the IEEE NLPKE 2011 (Natural Language Processing and Knowledge Engineering) in Tokushima, Japan.

$$p.K = \frac{|\{\, t \in p' \mid k(t) = 1\}|}{|\{t \in p' \mid v(t) = 1 \text{ or } a(t) = 1 \text{ or } k(t) = 1 \text{ or } og(t) = 1\}|}$$

$$p.OG = \frac{|\{\, t \in p' \mid og(t) = 1\}|}{|\{t \in p' \mid v(t) = 1 \text{ or } a(t) = 1 \text{ or } k(t) = 1 \text{ or } og(t) = 1\}|}$$

where *t* are the terms in *p'*, which is *p* modeled as a bag of words BOW [Gaussier and Yvon 2012], and *v(t)* (etc.) are the perceptual indicators of the term, which were already defined in the counter of the STEM algorithm (see Section 4.5) as

$$\textit{indicator level } p(t) = \begin{cases} 1 & \text{if stem(t) is in the lexicon of perceptual vocabulary} \\ 0 & \text{in all other cases.} \end{cases}$$

Profiles were not only calculated for each document (forum post), but also for each author by concatenating all posts of this author to one new pseudo-document and calculating its profile as described, which leads to another four-dimensional vector *vprofile(a) = [a.V, a.A, a.K, a.OG]*.

Similarity between text can be calculated in different ways [Gomaa and Fahmy 2013]. The similarity *vsim(p1,p2)* between two posts was measured as the cosine similarity of their profiles, which is defined as

$$\cos(x, y) = \frac{\langle x, y \rangle}{||x|| * ||y||}$$

The value cos*(x,y)*=1 indicates complete similarity, i.e. y=ax Ɛ R, whereas the value cos*(x,y)*=0 expresses orthogonality between the vectors x and y [Bayardo, Ma and Srikant 2007].

4.6.2 Hypotheses

Following the findings on cognitive style (see section 2.1.2), we think that every user has one or two preferred sensory systems that influence his or her choice of expressions. In this analysis step, we are not (yet) trying to prove the relation between a preferred sensory system and the use of sensory expression. Being the first to investigate this topic, we instead focus on the use of sensory expression itself in order to develop a sense of the potential of such an approach.

To evaluate whether such a preference concerning the use of sensory vocabulary exists, three assumptions need to be made and will be tested with our hypotheses:

- The choice of expression stays stable as long as the discussed topics remain within the same field. (One can expect that, e.g., in a discussion about photography there is an overall higher use of visual expressions than in a discussion about cooking. To address this problem, we chose different forums on different topics.)
- Individuals have different preferences regarding their choice of sensory expression, there is no complete dominance of one sensory system over all others (as sometimes stated, e.g., under the heading of the "age of visuality"). Also within a certain pattern of preferences, there may be differences concerning the intensity of the preference.
- The use of sensory vocabulary cannot be referred directly to the use of overall vocabulary.

These three assumptions may be combined in the following statement: Variance in the use of sensory expression is higher between texts written by different users than between texts written by the same user. This assumption might be altered by a tendency to adjust one's personal style of expression to the conversational partner. Both options cannot be explained only with the use of overall vocabulary. We propose to test this idea with 3 hypotheses which examine different parameters:

Hypothesis A states that every user has a preference profile for sensory modalities, expressed as a profile of usage of sensory vocabulary. Hence, posts written by one author should be more similar to one another than to posts written by somebody else. Of course, this will not hold for every two such pairs of posts, but only on average. We therefore posit as Hypothesis A:

$$avg_{p1,\, p2\, \epsilon\, P,\, p1.author=p2.author}\ vsim(p1,p2) >$$
$$avg_{p1,\, p2\, \epsilon\, P,\, p1.author \neq p2.author}\ vsim(p1,p2)$$

where the p_i are posts, P is the set of all posts, $p_i.author$ is the post's author, and *vsim* is the similarity between the VAKOG profiles of its two arguments. We restrict the comparison to those posts that have a distinguishable VAKOG profile in the first place, operationalized as a total number of VAKOG words larger than or equal to 3.

Hypothesis B states that the answers to a post p1 tend to have the same distribution of sensory vocabulary as p1. Again, we expect this to hold on average, restrict attention to posts with at least 3 VAKOG words, and state with a notation analogous to that in (3):

$$avg_{p1,\, p2 \in P,\, p2.refersTo = p1}\ vsim(p1,p2) >$$
$$avg_{p1,\, p2 \in P,\, p2.refersTo \neq p1}\ vsim(p1,p2)$$

Hypothesis C varies Hypothesis B. It starts from the assumption that an author will feel "attracted" to answer a post if this post matches his/her general sensory preferences, but that (for various reasons) the produced answer post may deviate from this general pattern. This leads to the expectation that the answering author's profile (rather than the answer post's profile in Hypothesis B) will correspond to the answered post. We restrict attention to those authors who have a distinguishable profile in the first place, operationalized as a total number of posts larger than or equal to 5:

$$avg_{p1,\, p2 \in P,\, p2.refersTo = p1}\ vsim(p1,p2.author) >$$
$$avg_{p1,\, p2 \in P,\, p2.refersTo \neq p1}\ vsim(p1,p2.author)$$

An overview of the 3 hypotheses and their methods of testing is given in Figure 24.

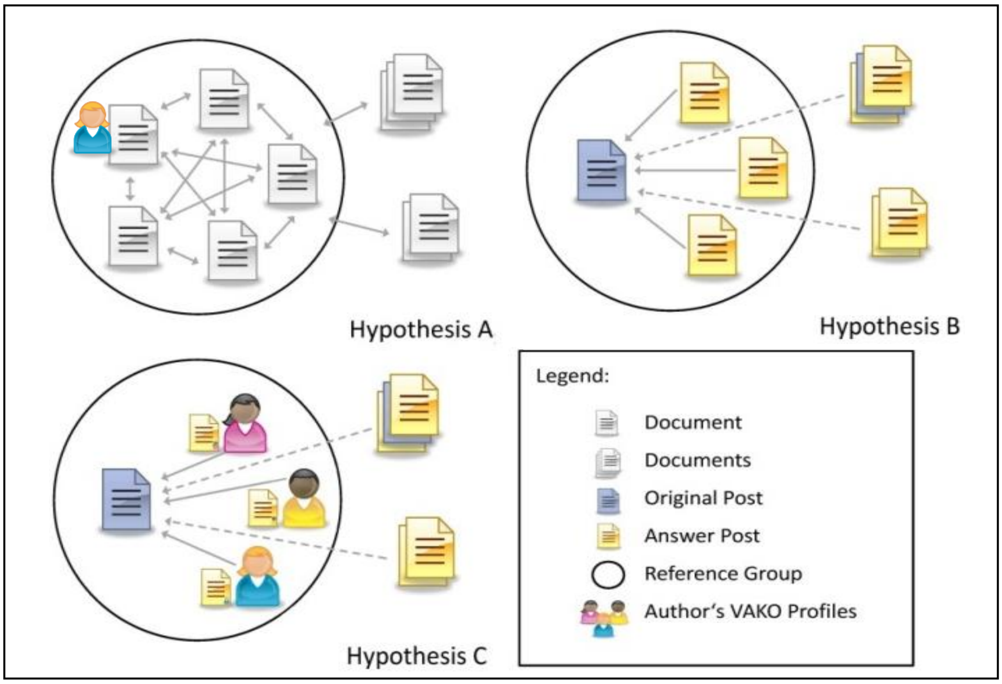

FIGURE 24 HYPOTHESES A–C AND THEIR TESTING

4.6.3 Data Preparation and Analysis Procedure(s)

The following four steps were executed in order to obtain analyzable data:

(1) We used Unix shell and PHP scripts to store the forum corpus, initially plain text with an xml-like tagging structure, in a MySQL database.

(2) We implemented the STEM algorithm for identifying and counting perceptual vocabulary in text with PHP. Each word of each post text within the corpus is hence initially looked up in the tagged terms database and, if not found, it is recursively reduced by one letter and looked up again. As soon as a match is found, the surrounding "leftover" letters are looked up in a German dictionary of all word forms and in several word component lists. If all letters around an identified sensory stem can be matched with existing components in compliance with the rules of German word formation, the counter of the respective perceptual system for the text is augmented by 1. The sensory profiles thus obtained for each forum post *vprofile(p)* are based only on the post body and do not take post headers into account. This was done because headers very often repeat the answered post's header and can therefore not be considered as original text written by the author of the post. For the same reason, we excluded empty posts from further analysis.

(3) The sensory profiles were then split into different subsets according to the needs of each hypothesis. For the subsets of hypotheses B and C, an answer tree structure was reconstructed with help of a PHP script running recursively from one post to its predecessor, following the reference number of a post. Every post was tagged as either being an initial post, a direct answer to an initial post or an answer to an answer post. In cases where referenced posts were missing (which might be the case for messages deleted by users or forum administrators), the post was handled as initial post.

The MySQL corpus database is at that point fully filled. Each forum post is stored as one record containing the following information:

- *Post ID:* a unique number of the post
- *Reference ID:* If the post is an answer to another post, the ID of that other post is stored in this field. Else if the post is an initial post (or if the reference ID points to a post that has been deleted and is hence not part of the dataset) the reference ID is the same as the post ID.

- *Initial Reference ID:* This field contains the ID of the initial post of the complete answer tree. It might be the same as the reference ID (if that one is an initial post already) and perhaps even of the post ID (if the post itself is an initial post).
- *Date:* the date when the post was published
- *Author:* the author's nickname
- *Title:* the post's title
- *Post:* the post's text
- *Letter count:* the number of letters in the post
- *Word count:* the number of terms in the post
- *v:* the number of visual terms in the post
- *a:* the number of auditory terms in the post
- *k:* the number of kinesthetic terms in the post
- *og:* the number of olfactory and gustatory terms in the post

For easier processing, a second table is built upon the first one, where each forum author is stored as one record containing the following information:

- *Author:* the author's nickname
- *Post count:* the number of posts written by the author
- *Letter count:* the number of letters in all posts written by the author
- *Word count:* the number of terms in all posts written by the author
- *v:* the number of visual terms in all posts written by the author
- *a:* the number of auditory terms in all posts written by the author
- *k:* the number of kinesthetic terms in all posts written by the author
- *og:* the number of olfactory and gustatory terms in all posts written by the author

Several constraints were applied when creating subsets: Whenever comparing one *vprofile* (of an author or a post) to another, all profiles with fewer than 3 identified perceptual terms were excluded. When looking at author profiles, only authors with at least 5 non-empty posts were taken into account. An overview of the reduction steps necessary for each subset is shown in Table 11. In view of the huge number of pairs in the complete set ("ALL") and the comparatively small percentage of pairs within the hypotheses sets, we chose to work with the same comparison set (including all pairs) for all three hypotheses instead of using a different comparison set (including all other pairs) for each hypothesis.

	Reduction criteria
ALL	• number of VAKOG terms (n_v) per post>2
Hypothesis A	• min. 2 posts with n_v>2 per author
Hypothesis B	• min. 1 initial post and 1 corresponding answer post with VAKOG-sum per post >2
Hypothesis C	• min. 1 initial post with VAKOG-sum >2 • 1 corresponding answer post from an author with min. not-empty posts per author >4

TABLE 11 REDUCTION CRITERIA PER SUBSET

(4) All hypotheses were tested in two ways. First, a comparison of means was used to test the hypothesis directly. However, the result of such a comparison may yield a larger VAKOG similarity of, for example, a post and an answer to it, simply because these two documents will discuss similar content and therefore probably also use similar language. We decided to control for this by treating the full-text similarity of the two posts as a covariate. The hypotheses then are refined to "if two pairs of posts each have the same full-text similarity, the pair in which one answers the other will have higher VAKOG similarity than the unrelated pair" (analogously for the other hypotheses). Full-text similarity was operationalized as the cosine similarity between the two posts modeled as bags of words (BOW) by the WEKA[13] StringToWordVector filter, using standard parameters: word counts of a maximum of 100,000 words weighted by TF.IDF, which calculates the importance resp. relevance of a word to a document by comparing its occurrence in the specific document to its overall occurrence.

Since similarities were not normally distributed, we opted for a non-parametric test of these statements, choosing loglinear modeling for 3-way contingency tables. The values of the three dimensions were: (i) pairwise full-text similarity, binned into *n* equal-sized intervals (in our case of *n=10: [0, 0.1), [0.1, 0.2),* etc.), (ii) pairwise VAKOG similarity, binned into the same *n* intervals, and (iii) the 2 categorical values of the variable of interest (e.g. *answer-post-relation* vs. *non-answer-post-relation*).

The statistical analysis was done in Excel and with the help of an online tool for the comparison of 3-way-contingency-tables[14].

[13] http://www.cs.waikato.ac.nz/ml/weka

[14] http://faculty.vassar.edu/lowry/abc.html

4.7 The Vocabulary of Perception in Forums[15]

The following chapter explores the use of the vocabulary of perception in forums. To ground further investigation in the relation between the use of sensory vocabulary and sensory preference, several basic questions about sensory word use in forums need to be answered: a) Is sensory vocabulary used sufficiently often to be considered as an indicator for individual preference? b) Does the use of sensory vocabulary stay consistent within the posts of one user? c) How can the use of sensory vocabulary be modeled?

Therefore, a corpus with more than 1.000.000 forum posts was analyzed for the occurrence of expressions that are directly linked to a sensory system. We found that users differ significantly in their use of sensory expressions and that most users have preferred patterns for the use of sensory expressions. Furthermore, we found a correlation between the sensory vocabulary of a post and the sensory preference of the users who answered this post.

4.7.1 E30 Forum Corpus

To test the hypotheses presented in section 4.6.2, we chose Richling's forum corpus [Richling 2008], which can be accessed freely after registration via web interface.[16] It is a corpus built on posts from discussion forums on the car type BMW E30, published in the years 2000 until 2007. This very narrow topic helps to exclude result variation due to discussion of different topics. The corpus is monolingual and consists of more than one million posts in German, each post text accompanied by information on the author, the header, the reference post, and the date.

All data was prepared according to section 4.5. An overview of the reduction steps necessary for each subset and the resulting amount of data is shown in Table 12.

[15] This section is based on the publications „Towards a New Dimension for User Modeling: The Use of Sensory Vocabulary“ [Kellner and Berendt 2012], which was awarded with the „Best Poster Award“ at the UMAP 2011 (User Modeling, Adaptation and Personalization) in Girona, Spain; and on its extended version „Extracting Knowledge about Cognitive Style. The Use of Sensory Vocabulary in Forums: A Text Mining Approach“ [Kellner and Berendt 2011], which was awarded with the „Best Paper Award“ at the IEEE NLPKE 2011 (Natural Language Processing and Knowledge Engineering) in Tokushima, Japan.
[16] http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/institutkorpora/

	Reduction criteria	Posts left for further analysis	Pairs of posts left for further analysis
ALL	• number of VAKOG terms (n_v) per post>2	97.951	4.797.150.225
Hypothesis A	• min. 2 posts with n_v>2 per author	54.182 posts of 5.879 authors	9.587.622
Hypothesis B	• min. 1 initial post and 1 corresponding answer post with VAKOG-sum per post >2	34.419	22.796
Hypothesis C	• min. 1 initial post with VAKOG-sum >2 • 1 corresponding answer post from an author with min. not-empty posts per author >4	187.273	158.885

TABLE 12 E30 FORUM CORPUS: REDUCTION CRITERIA AND THE RESULTING AMOUNT OF DATA PER SUBSET

4.7.2 Results

The E30 forum corpus consists of 1,053,841 posts, written by 30,021 different authors. The corpus contains 26,305,285 terms (defined as strings of characters and/or numbers, separated by blanks and/or punctuation marks).

We categorized 21,037 different terms as sensory expression, 6,798 of them as visual, 5,047 as auditory, 7,518 as kinesthetic and 1,674 as olfactory and gustatory and stored them in a dictionary. Naturally, most of the categorized terms appeared more than once in the corpus. The corpus contains 785,303 terms classified as sensory expression, 318,305 of it visual, 248,896 auditory, 192,566 kinesthetic and 25,536 olfactory and gustatory terms. 336,055 posts (52.0% of the total amount of non-empty posts) contained at least 1 sensory term; 97,951 posts (15.2% of the total amount of non-empty posts) contained at least 3 sensory terms which was used as filter criterion when comparing single posts. Looking at the distribution of sensory expression within the E30 forum corpus, around one out of 30 terms could be classified as sensory expression. (By manual analysis we found that there is still a large number of non-systematic "dummy terms" within the corpus replacing original links, images, references to other postings etc. The occurrence of sensory expression might thus be higher in text corpora containing only "real" words). Even though the used number of sensory terms might not seem very high, it is sufficient to obtain relevant information on the average user (who publishes 35 posts). A detailed distribution can be found in Table 13.

Posts	1,053,841
Original posts	223,973
Answer posts	829,868
Not-empty posts (neP)	646,455
Av. nr of terms per neP	40.69
Authors	30,021
Av. nr of posts per author	35.10
E30 Dictionary[17]	
Different terms	474,264
Different visual terms	6,798
Different auditory terms	5,047
Different kinesthetic terms	7,518
Different olfactory+ gustatory terms	1,674
Sensory terms	785,303
Visual terms	318,305
Auditory terms	248,896
Kinesthetic terms	192,566
Olfactory+gustatory terms	25,536

TABLE 13 THE DISTRIBUTION OF SENSORY EXPRESSION IN THE E30 CORPUS

Concerning our hypotheses, we applied two methods of testing.

(1) Comparison of distributions with Mann-Whitney's U test: Values for each set and for both full-text and VAKOG similarity distribution were calculated. Significance testing against the null hypothesis of equal distribution was calculated separately for full-text and for VAKOG similarity and for each hypothesis, comparing the hypothesis set with the set of all post pairs. The results can be found in Figure 25. All comparisons were statistically relevant with a p-value<.0001.

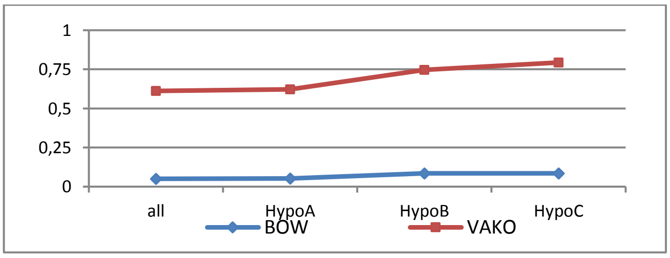

FIGURE 25 FULL-TEXT AND VAKOG SIMILARITY IN OUR SAMPLES

(2) To combine the full-text and VAKOG similarities, we used loglinear modeling for 3-way contingency tables. Results were calculated comparing all 3 tables at once and comparing pairs of 2 tables collapsed across the levels of the third. The results can be found in Figure 26. All comparisons were statistically significant with a p-value <.0001.

[17] Terms are handled as different as soon as they differ in one letter, including versions due to typos.

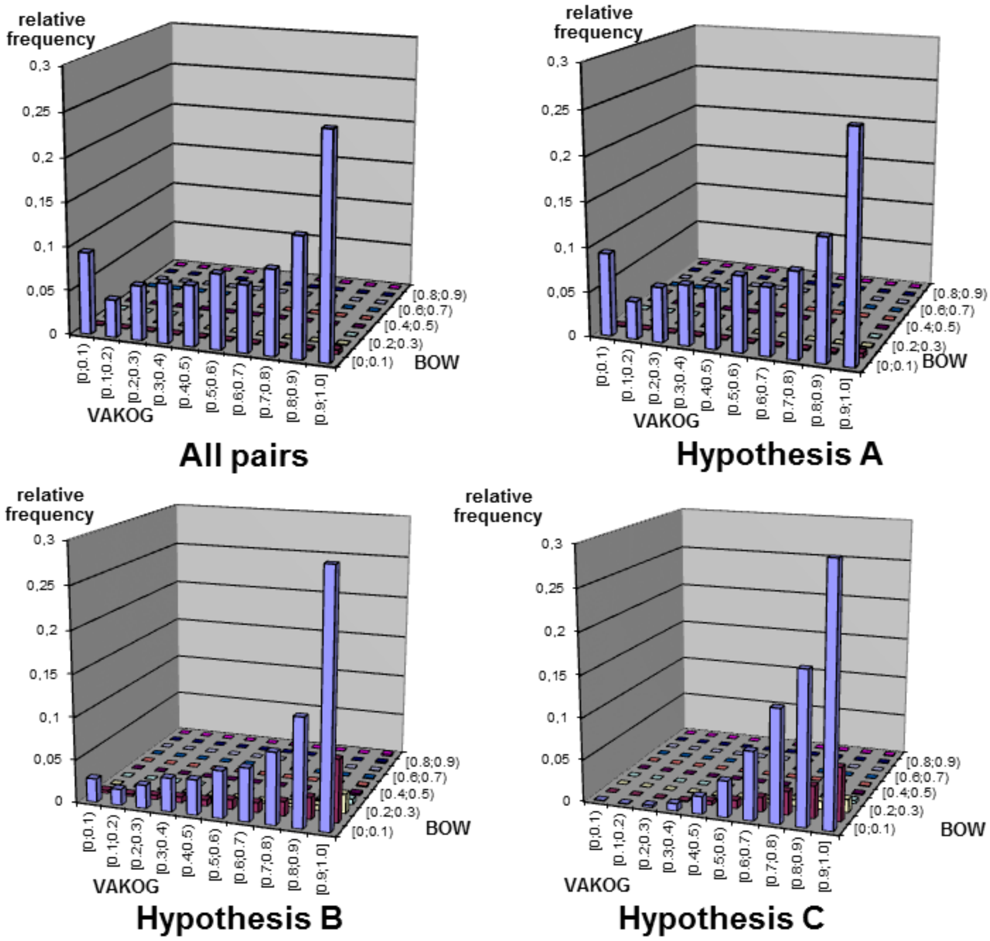

FIGURE 26 DISTRIBUTION OF FULL-TEXT AND VAKOG SIMILARITY WITHIN THE SAMPLES

Overall, the hypotheses testing on the E30 forum corpus shows a significantly higher VAKOG similarity within the hypotheses subsets than within the set of all post pairs. These results were consistent both for the comparison of means and the combined full-text and VAKOG similarities. That confirms all three hypotheses and leads to the following conclusions: Authors of forum posts have a tendency to use sensory expressions in similar distributions over time and to rather answer posts corresponding to their individual distribution patterns. Also, distribution patterns show a higher similarity within answer threads.

CHAPTER 5

GETOLS: Game Embedded Testing of Learning Strategies[18]

To see clearly, it often needs only a change of perspective.
(Antoine de Saint-Exupéry)

The third test approach developed within this thesis is an embedded testing method. Learning strategies are typically tested in rather conventional test settings, using questionnaires or observation as a method of information acquisition. However, such conventional test settings are far from the test taker's everyday life situation and might therefore evoke stress. Embedding a learning strategies test into an educational computer game, provides several advantages over the conventional testing procedure: The test takers can stay in their familiar learning environment and even learn something during the process of being tested. We conducted an explorative study by testing 24 pupils, comparing the outcome of a conventional learning strategies test conducted by a psychologist with the results obtained by our GETOLS method. The high similarity of results strongly supports our approach.

[18] This chapter is based on the publication „GETOLS – Game Embedded Testing of Learning Strategies“ [Kellner and Weißenbacher 2012], which was awarded with the „Best Student Paper Award“ at the IEEE DIGITEL 2012 (Digital Game and Intelligent Toy Enhanced Learning) in Takamatsu, Japan.

5.1 Introduction

According to Prensky [2001], the main idea of digital game-based learning is to connect learning content and computer games in order to achieve an equal or better learning success than with traditional learning methods. Prensky defines digital game-based learning as "[…] any learning game on a computer or online", under the condition that engagement and learning are always kept at a high level. A strong connection of learning and play is very important, which means that game elements and learning should not fall apart as it was the case in most edutainment titles.

Advantages of game-based learning are, amongst other: the added engagement, the interactive learning process employed and the way they are put together. It has been proven that educational games significantly enhanced student learning relative to nongame conditions [Clark, Tanner-Smith and Killingsworth 2016]. In addition, computer games can serve as appropriate learning objects for the digital natives generation [Pivec 2008] and can build upon well-established learning principles [Clark and Mayer 2016]. According to Kruse [2002], the following specific advantages of e-learning can be identified among others: reduced overall costs, increased retention, on-demand availability, self-pacing, interactivity etc. Furthermore, educational games have been proven to be a source of enjoyment [Iten and Petko 2016] and, if well-made, to evoke a flow feeling within the players [Kiili 2005b]. If games are challenging enough, the engagement as well as the learning outcome are higher [Hamari et al. 2016]. However, there are also some concerns about digital game-based learning: The main objection is that it is critical to convey that learning has to be fun whereby the learning itself is devaluated [Okan 2003]. If learning is not properly integrated into the gameplay the main amount of time is spent on playing and not on learning [Egenfeldt-Nielsen 2007].

Altogether, educational games are a very promising tool for learning under the conditions that they are well-made and employed in a suitable way. They can be used in different contexts and by different people, in vocational education and training, in schools and for informal learning. Some games aim to teach special learning contents (from contemporary history [Šisler 2016] to art education [Froschauer et al. 2011]), others intent to train skills [Ju and Wagner 1997; Romero, Usart and Ott 2014] or try to trigger the players' reflection on social topics like alcohol and drug abuse

[Rodriguez, Teesson and Newton 2014], political efficiency [Huang Ling 2011] or sustainable behavior [Mercer et al. 2017]. Teachers are using educational games in schools, children and students play educational games in their spare time, enterprises as well as the military use educational games for training purposes.

At the same time, taking a psychological test is often perceived as a rather stressful situation [Sarason 1984]. By embedding a psychological test into an educational game, the aim is to combine the advantages of gameplay with the outcome of a learning strategies test. In order to examine this approach, we compared the results achieved via GETOLS method with the ones collected by a psychologist.

In this chapter, we face the following questions: a) Is it possible to embed a psychological test into a game? b) How could a psychological test be adjusted in order to make it feel game-like? c) What are the benefits of game embedded testing?

The chapter is organized as follows: Section 2 searches for existing answers to those questions in related research. Section 3 explains the methods we used to find our own answers to those questions. Section 4 presents the results, which are discussed in section 5. Section 6 provides a conclusion and an outlook to further research.

5.2 Related Research

An overview of cognitive styles, especially those that reflect upon sensory preference, has been given in section 2.1. In this section, we explore related research specifically needed for this chapter. We start with a section on digital game-based learning. Second, we take a look at test theory and psychological testing methods with a focus on personality or preferences. Third, we focus on the field of game embedded testing.

5.2.1 How to Test Learning Strategies

As described for the field of user modeling, also in the field of psychological research, two sorts of variables are tested in order to describe a person: rather permanent and rather transitory attributes. Learning strategies can be categorized as rather permanent attributes. They are developed according to conscious or unconscious interventions by the learner. Furthermore, interventions might be formed by some other external agent [Robotham 1999]. Learning strategies might change after a longer period of time or after intense training, but not from one day to the other [Reid 1987].

Even though learning strategies tests, like most other psychological tests, are traditionally done face to face or via questionnaire, a lot of research has been invested during the last decades into the topic of shifting the testing procedure to computers. If well implemented, computer based testing can be treated as equivalent to pen-and-paper testing [Bugbee 1996].

5.2.2 Embedded Testing

According to Aleksić and Ivanović [2017], a high percentage of empirical research focused on the effects of digital games found correlations between playing digital games and learning styles. Testing the learner’s learning achievement and learning gains can be done in explicit and implicit manner, the assessment should be well integrated into the game [Bellotti et al. 2013].

Explicit testing of learning strategies can be evaluated with the help of questionnaires or explicit tasks that need to be fulfilled, and is performed to test the learner’s knowledge. Educational games are designed in order to enhance the player’s knowledge about a certain topic. Naturally, that knowledge should be tested during or after the game. On the one hand, such testing of knowledge works as repetition and supports learning efficiency; on the other hand it might be used to prove teaching efficiency [Morris and Fritz 2000]. Ramani et al. [2008] provide a skin for explicit knowledge testing within a game. Such knowledge gain can be analyzed along match or mismatch of the individual learning strategies.

Implicit testing of learning strategies is a rather new approach in the field of digital learning that analyzes the learner’s behavior during the learning process, and rather focuses on the learner’s learning approaches and skills. What can be tested and how that can be done, is object of investigation in the field of embedded assessment. Underwood, Kruse and Jakl [2010] describe embedded assessment as “the process of measuring knowledge and ability as part of a learning activity”. Subject of examination are, e.g., problem-solving strategies [Ketelhut, Nelson and Schifter 2009] and higher-order skills [Shute et al. 2008]. Feldman, Monteserin and Amandi [2014] deduce learning style from the learners’ in-game-behavior in a puzzle game.

Our approach applies mainly to the first group: Even though the testing phases are well embedded in the storyline and thus somehow “hidden” in the game, the testing itself is explicit in order to keep the results comparable to the ones collected in the conventional test setting.

5.3 EAGC: An Excursus to Educational Game Design[19]

When designing the game, we found there was not enough helpful guidelines to do so. Despite the recent popularity of game-based learning and some first general guidelines for the creation of such educational games [Moser 2000; Kiili 2005a; Quinn 2005; Amory 2007], there is a lack of useful practical guidelines for specific game types that address all relevant aspects of design, implementation and testing. This might be explained with the possible lack of experience of instructional designers with computer games and game designers with education [Hirumi et al. 2010], but it influences the focus and the approach chosen for the game design and implementation process. Therefore, we developed the EACG guidelines for the creation of an educational adventure game. The goal of our work was to develop guidelines for the creation of educational adventure games that help not to forget any aspect. In order to do so, we refer to both existing guidelines for the design of entertainment games and existing frameworks for the design of educational games.

For planning and implementing an appealing educational game, it is essential that the issue of game design is considered. What makes good game design? To answer this question is hard, because there are no general rules. A variety of studies present different aspects that should be considered. Bond and Beale [2009] identify four important key elements for good game design by examination of computer game reviews: variety, cohesion, social and user interaction. Prensky [2001] stresses six elements of successful games: good game design is balanced, creative, focused and

[19] In this section, in opposite to the rest of this chapter, “we“ refers to the cooperation between the author and Paul Sommeregger, who wrote his master thesis under the author’s co-supervision. This section is based on the publication “Guidelines for Educational Adventure Games Creation (EAGC)“ [Sommeregger and Kellner 2012], which was written for the IEEE DIGITEL 2012 (Digital Game and Intelligent Toy Enhanced Learning) in Takamatsu, Japan, where we also presented our findings on GETOLS.

has character, tension and energy. Schell [2015] summarizes a hundred guidelines which should be taken into account in order to create an appealing game.

Summing up the work of Schell [2015] and Adams [2014b], the following important elements of computer games can be identified:

- Game mechanics: The game mechanics are the fundamental component of every computer game. Game mechanics describe all the procedures and rules of the game and determine its challenges and goals and how these can be achieved or overcome.
- User interface: The user interface is the link between the player and the game mechanics. It presents the game challenges to the player and transforms the input of the player into actions in the game.
- Gameplay: The gameplay consists of the challenges presented to the player and the actions available to master them (jump, run, interact, etc.).
- Story: Not every computer game has to have a story but for the genre of adventure games this element is very important. In short, a good story is a report about a sequence of events which has to be at least credible, coherent and dramatically meaningful. In a computer game, stories should be interactive so that they can be influenced by actions of the player.
- Characters: An integral ingredient of the story are the characters, which breathe life into the game world and make a major contribution to a compelling game experience. It is important that their appearance and their behavior is consistent. Another crucial point is that they should have a certain value of recognition and that the player can identify with them. The latter is especially vital for the avatar (the character which is controlled by the player).
- Game world: The game world is the location where the events of the game take place. In most games it is represented with the help of auditory and visual media.
- Puzzles/Challenges: An essential element of every computer game are challenges and puzzles. It is important to make sure that puzzles are clearly presented to the player, so that he or she knows what to do and that the player is provided with a certain feeling of progress dealing with the puzzle to avoid frustration. In order to evoke such a feeling of progress, the puzzles should be constructed incrementally in difficulty. Sometimes it can be necessary to provide the player with hints in order to avoid stagnation.

All these elements should support learning. Tan, Ling and Ting [2007] divide these components into pedagogical components (difficulty to learn, exploration, challenge, engagement, goal setting etc.) and game design components (interaction, storytelling, interface, simulation, feedback etc.). They studied several design frameworks for educational games and identified three major subcomponents which educational game design should contain: multimodal (modality and interaction), task and feedback. Learning instructions and feedback foster deeper learning [Erhel and Jamet 2013].

The design process of different game genres might demand different design approaches [Kiili 2005a]. Setting an emphasis on adventure games, we suggest a structure of five main game development phases, namely conceptual design and game design, implementation, testing and validation. We also take project management into account, to not only guide through the creation of the game itself but also to support the organization of the game development process.

Those five main game development phases are:

5.3.1 Conceptual Design

The first stage is the stage of conceptual design, in which the basic decisions for the creation of the game are made. As Adams [2014b] stated in his model for game design, the outcome of this stage can hardly be changed later on. Since we want to represent guidelines for educational games creation, we have to start where every good instructional design has to start: First, the topic and the target audience must be defined. This is a mutual process because not every topic is suitable for every target audience and the other way round. It is also important to choose a topic that is suitable for implementation in an adventure game. According to Moser [2000], topics that only consist of “facts rather than rules for their interrelation may not be appropriate”. When a suitable topic is chosen, the target audience needs to be analyzed regarding their interests, motivation, existing knowledge and possible misconceptions [Quinn 2005].

With the topic and the target audience along with their characteristics defined, the topic needs to be searched for appropriate learning goals, and a suitable methodical setting should be defined. The appropriate learning goals and the methodical setting are once again influenced by each other. Methodical setting means the applied learning theories (constructivism, behaviorism, cognitivism), the level of difficulty and how assessment and adaptivity should be integrated. Constructivism with its concepts of situated learning and discovery learning seems to be especially suitable for didactic adventure games. The level of difficulty is determined by the target audience and is connected to adaptivity. In order to integrate adaptivity into the game, assessment should be embedded into the game environment. Embedded assessment can be used to measure all types of knowledge (content, process, skills) that can be achieved throughout playing an educational game [Underwood, Kruse and Jakl 2010].

The next step is to choose the relevant learning goals out of the range of identified possible learning goals. This means to select the learning goals the player should achieve by playing the game. Subsequently, corresponding units of meaning and sub-goals for the chosen game objectives need to be defined. These units of meaning and sub-goals will form the basis for the creation of puzzles in the next design stage.

5.3.2 Game Design

Building upon the specifications made in the conceptual design stage, the main game creation process begins and details are shaped [Adams 2014b].

First of all, the game setting (which depends on the target audience), the methodical setting and the learning goals need to be determined. It is important that the player can identify with the setting in order to get engaged [Moser 2000]. This means that the setting needs to be adapted to the specific target audience. At this point, gender should be considered in order to create a game that is appealing for both sexes. Primarily, stereotypes should be avoided [Mou and Peng 2009]. The setting must also be matched to the learning goals and the methodical setting in order to make their seamless integration possible. Quinn [2005] states that the theme of an educational game should be more concrete if a specific skill should be learned and more fanciful if the skill is more general.

The next step is to develop the storyline along with its characters, the game world, and the puzzles related to it. Those parts can, to a large extent, be developed simultaneously.

Based on the units of meaning and sub-goals, according puzzles need to be created and embedded into the storyline and the game world. This process should make use of the strong connection of place and content of adventure games.

Following the suggestion of Moser [2000], the game world should be divided into several sections. Every room of a section tackles a specific unit of meaning or sub-goal. The player needs to get involved with these units of meaning and sub-goals in order to achieve the specific learning goal related to them and to proceed to the next section. A transition from one section to another should also be connected with a climax in the story of the game. Amory [2003] suggests that every act of the story involves several sections. The rooms in every section are freely accessible in order to provide a feeling of freedom to the player and to provide discovery learning. The final puzzle should be a task where all the acquired knowledge has to be applied.

As far as puzzles are concerned, one should integrate pedagogical support, feedback, and treatment of potential misconceptions in order to avoid frustration [Burgers et al. 2015]. Again, they need to be implemented into the game world and the storyline in order to maintain the gaming experience. Such support could be hidden in books with hints, advises from characters or other built-in support.

Another important aspect connected to the game setting is the user interface. In adventure games, this is mainly a point and click interface, which means that the user manipulates things on the screen by clicking on them with the mouse cursor. This user interface and game play mode is extremely suitable for discovery learning though other types of user interfaces are conceivable. Specific puzzles could, for instance, demand other game play modes and subsequently the definition of other user interfaces. Nonetheless, every user interface should be designed in a way that it is appropriate to the setting.

5.3.3 Implementation

If the game design stage has reached a point at which the implementation of the conceived aspects seems to be necessary, the implementation stage begins.

First of all, a programming environment needs to be chosen. In the field of adventure games, there are a lot of free programming environments available like OpenSludge[20], AGS[21], Wintermute Engine[22] or, especially for educational games, the <e-Adventure> platform[23].

Once this choice is made, programming can start and game mechanics are established. This process is guided by the creation of the media for the game. Graphical design, sound design and music design must be approached with the target audience and setting of the game in mind. Voice output has to be created apposite to the game characters.

5.3.4 Testing

After the implementation stage has been completed and a playable version of the game has been created, the game needs to be tested for several aspects, like functionality, usability, storyline integrity and difficulty. Functionality has to be verified in order to avoid gameplay errors. Test for usability means, e.g., to test whether the user can handle the user interface effectively [Isbister and Schaffer 2008]. This can be tested by defining tasks the user can accomplish through the use of the user interface [Quinn 2005]. Storyline integrity needs to be validated in order to avoid logical errors or goofs. Finally, the level of difficulty needs to be tested to ensure that it is appropriately balanced [Hamari et al. 2016]. The so-gained feedback should be used to identify improvements for the game and its design. At that point, the design process is taken back to the game design stage in order to implement the improvements.

5.3.5 Validation

Once the game design process is finished, the validation stage begins. In this stage, educational effectiveness and engagement are tested with the target audience. Tests for educational effectiveness refer to whether the learning goals have been achieved or not [All, Nuñez Castellar and Van Looy 2016]. Learning assessment comprises the learning progress as well as the outcomes [Bellotti et al. 2013]. A test for engagement

[20] http://www.opensludge.github.io
[21] http://www.adventuregamestudio.co.uk
[22] http://dead-code.org
[23] http://e-adventure.e-ucm.es

is more complex as engagement is a rather subjective term. It can, for instance, be captured by the degree that someone liked the game or not, or by the flow feeling evoked [Kiili et al. 2014]. The results of this final development stage can once again be taken into account in order to improve and refine the game until a satisfying product has been created.

5.3.6 Project Management

As we have seen, the creation of an educational game is a very complex task. In order to master it successfully, terms of project management need to be regarded as well. Based upon literature about the topic of project management [Collins 2011; Heagney 2011] and game design [Iuppa and Borst 2010; Adams 2014b], we will give a brief overview about the most important steps of project management as far as the creation of educational games is concerned.

Project Management is divided into four major phases: project initiation, project planning, project execution and project closure.

In the project initiation phase, first of all, the treated problems and goals of the project need to be defined. For the creation of an educational game, this means to determine the reasons for the creation of the game and the goals which should be achieved with it on a very high level. This includes determining the target audience and the addressed topic. In the next step, potential solutions are analyzed in order to get an idea of the overall theme of the game. After a certain solution has been identified, the project has to be shaped in terms of vision, objectives, scope and deliverables. Roles and responsibilities need to be determined and a rough project plan including the necessary funds needs to be established, which also includes the treatment of potential risks and opportunities for the project. This is followed by the appointment of the team in order to integrate all members of the team as soon as possible into the creation process. Next, a rough description of the game should be developed, the so called “high concept statement”. After that, a game treatment document needs to be created, which describes the game more in detail. Both documents, the high concept statement and the game treatment document, are sales tools that should help to raise funds and generate interest for the game.

As soon as the rough project planning has been managed and the project is funded, the planning phase starts. This means the creation of a detailed project plan, resource plan and financial plan including milestones, tasks, and resources needed. In addition, a detailed project schedule is formed to assess the progress of the project.

After the project has been planned successfully, the execution phase begins. For an educational game, this means that the game elements (levels, story, characters, game world, and puzzles) are designed and implemented. To support this process, several documents are generated: the character design document which describes the created characters in detail, the world design document which forms the basis for the creation of all the artwork of the game, the flowboard to illustrate the different gameplay modes, the story and level progression document which documents the overall story and the progress from level to level, and the game script which specifies the rules of play [Adams 2014a]. These documents are very important in order to record decisions, to shape ideas and to communicate the design concept to other team members.

The execution phase is accompanied by monitoring and controlling, which encompasses for instance time management, cost management and quality management.

When the project has reached its completion, the closure phase begins, and the outcome of the project is analyzed and documented. In an educational game, such an analysis should also address educational effectiveness of the game and the players' engagement.

Bringing all this together, we designed a flowchart showing the most important steps for the design and creation of an educational adventure game. The complete EAGC process is modeled in Figure 27.

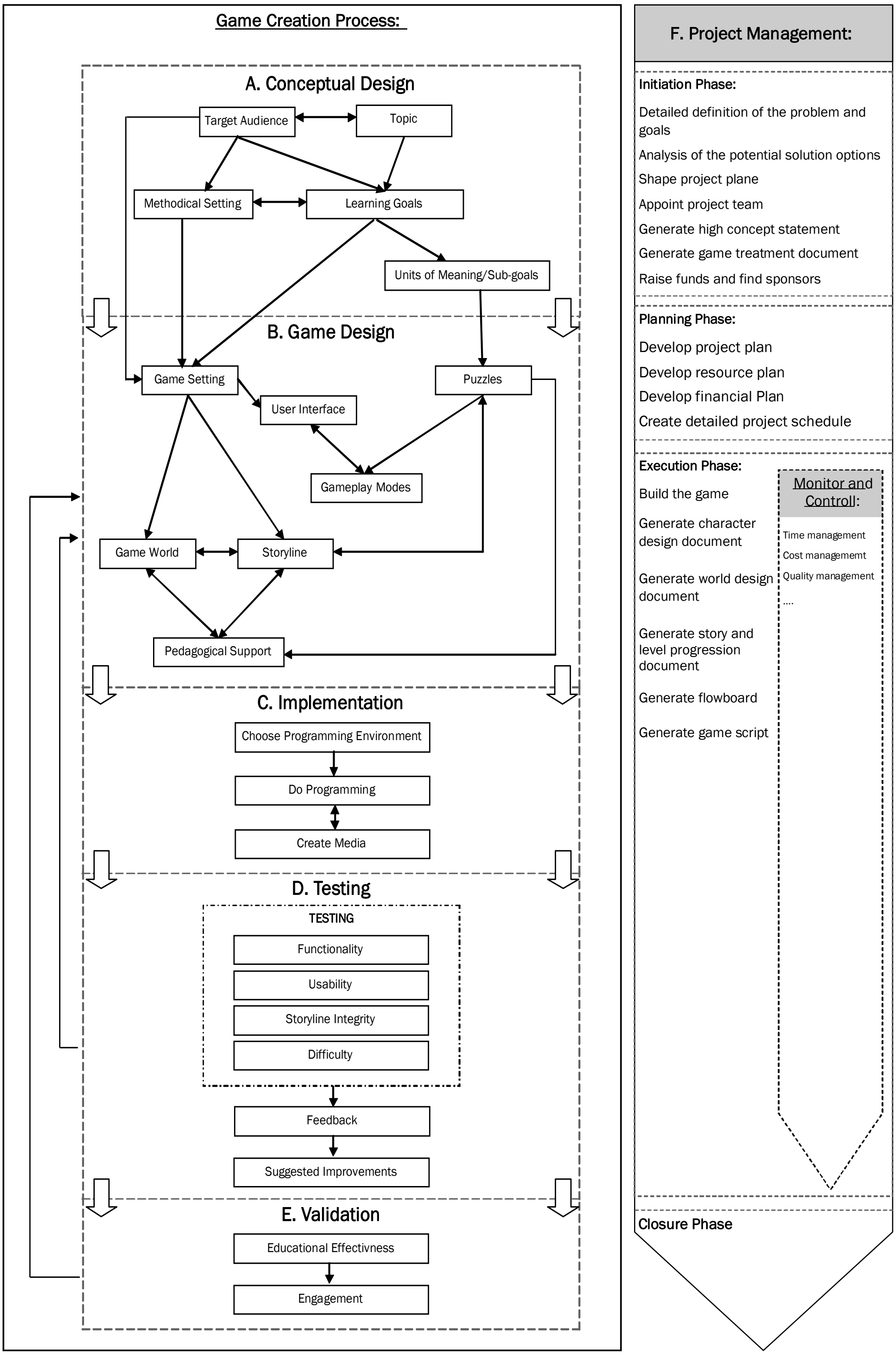

FIGURE 27 THE EAGC PROCESS: STAGES AND DEPENDENCIES

5.4 Method

Before defining the method, we want to sum up the goals of this research:

- to ease the testing process of learning strategies by embedding it into a game
- to choose the learning strategies model most adequate for implementation in a computer game
- to push the testing process into the background of the software in order to make the test takers feel comfortable and not like taking an exam
- to make the game (and hence the testing process) educational and enjoyable

5.4.1 Testing Method

In order to find the learning strategy test most suitable for application in GETOLS, we analyzed nine learning strategy models. We chose three models for more detailed analysis: Kolb [1984] presents in his learning style inventory the two axes concrete vs. abstract and reflective vs. active. The four quartiles are classified as four learning styles, namely converging (abstract, active), diverging (concrete, reflective), assimilating (abstract, reflective), and accommodating (concrete, active). Pask [1976] differentiates in his work two different learning approaches, serialists (partists) and holists (wholists). Vester [2011] describes in his model of learning strategies four perceptual preferences, namely visual, auditory, kinesthetic, and text based input.

For the classification of the models we address the following five categories of suitability:

- Technical applicability (TA): How complicated is the technical application of the test to a game?
- Game integratability (GI): Is it possible to integrate the theoretical concept into a game?
- Complexity of test evaluation (CTE): How complex is the test evaluation?
- Usefulness of results (UR): Are the results understandable and helpful for the test taker?
- Name recognition (NR): Is the concept well known to potential future users? In order to address a high number of users, the concept should not only be known among researchers, but also among people working in the educational sector.

Every category is assigned one value per model, choosing between the values + (very suitable), ~ (quite suitable), – (not suitable at all).

Kolb's testing method is based on a questionnaire. It is hence very suitable in means of technical application and complexity of test evaluation, but very hard to embed into a game without interrupting the story flow. The results rather inform about abstract strengths than about concrete learning strategies. The model is known to researchers and interested laypersons, but not to the general public.

Pask's testing method is also based on a questionnaire and hence shows the same characteristics as Kolb concerning technical applicability, game integratability and complexity of test evaluation. Its results inform about the individual information processing style. The model is known to researchers and interested laypersons, but not to the general public.

Vester's testing method is based on a memory test: for each testing sequence, ten objects are presented in different perceptual modalities (visual, auditory, kinesthetic and text based input) and shall be recalled by the test person after a distraction sequence. The testing is not so easy to realize in means of technical applicability, but it is more suitable concerning the game integratability. The test evaluation is more complex because of the necessity of free word input, which might need special treatment of spelling mistakes and the use of synonyms. The results provide explicit and individual suggestions for improvement on how to learn. The model is very well-known in the didactic sector in German-speaking countries.

An overview of this analysis can be found in Table 14.

Method	**Categories of Suitability**					**Sum**
	TA	***GI***	***CTE***	***UR***	***NR***	***+ = 3*** *~ = 1* *– = -3*
Kolb	+	–	+	~	~	5
Pask	+	–	+	~	~	5
Vester	~	~	~	+	+	9

TABLE 14 SUITABILITY OF DIFFERENT TESTING METHODS FOR GETOLS

According to this suitability analysis, we opt for the testing method of Vester. Even though the technical application is a bit more complicated and needs some further investigation in mapping to digital evaluation, the testing method is easier to integrate into a game. The interpretation of disambiguate input can be solved with more or less work dependent on the degree of automation. Out of the analyzed three models, it is the most popular one and is widely used in the educational sector. The most important point is the usefulness of results, which is assured by the detailed feedback provided to the test taker.

The mapping of object presentation to a digital format is quite easy for visual, auditory and text based modalities, but a challenge regarding the kinesthetic modality. According to Vester [2011], not only tactile input, but also the use of an object may be perceived as kinesthetic experience. We therefore produced videos showing the use of objects so as to evoke similar results. For validation of our mapping, we conducted a small study with 20 test persons, performing first the original test with direct kinesthetic input and then our GETOLS video test. For all test persons, the results only differed very little. The results of the study are presented in Table 15.

	Mode of Kinesthetic Testing		**Difference**
	Vester	***GETOLS Video***	
Subject 1	6	7	1
Subject 2	8	9	1
Subject 3	5	6	1
Subject 4	6	7	1
Subject 5	4	5	1
Subject 6	5	5	0
Subject 7	7	8	1
Subject 8	3	5	2
Subject 9	3	4	1
Subject 10	5	4	-1
Subject 11	4	5	1
Subject 12	7	6	-1
Subject 13	8	10	2
Subject 14	6	7	1
Subject 15	5	5	0
Subject 16	3	4	1
Subject 17	4	4	0
Subject 18	6	5	-1
Subject 19	7	7	0
Subject 20	6	8	2
Average⃠	**5.40**	**6.05**	**0.65**

TABLE 15 VALIDATION OF GETOLS VIDEO USE FOR KINESTHETIC TESTING

The scores for the testing method using GETOLS video are slightly higher than the scores for Vester's testing method, which might be explained with a learning effect from the first to the second testing scenario. Still, deviation is of 2 points at most, the average of 0.65 lies beneath one point. We consider the use of videos hence as an adequate mapping of the kinesthetic testing method to a digitally realizable version.

5.4.2 Measures

The outcome of Vester's learning strategies test are four values, pointing out the memory's performance in regard to information presented in four different ways, namely visual (V), auditory (A), kinesthetic (K) and text based (T) input. Each value is in the range of 0 and 10. The test results can be expressed as a four-dimensional vector

$$vresult(t) = [p.V, p.A, p.K, p.T]$$

based on the number of memorized objects per input system.

The vector similarity *vsim(t1,t2)* between two test results was measured as the cosine similarity of the two result vectors, which is defined as

$$\cos(x, y) = \frac{\langle x, y \rangle}{||x|| * ||y||}$$

The value cos*(x,y)*=1 indicates complete similarity, i.e. y=ax Ɛ R, whereas the value cos*(x,y)*=0 expresses orthogonality between the vectors x and y [Bayardo, Ma and Srikant 2007].

5.4.3 Hypotheses

We are testing the following three hypotheses:

Hypothesis A states that Vester's learning strategies test can also be performed digitally via the GETOLS method. This is the case if hypothesis B can be verified.

Hypothesis B investigates the similarity of the two testing methods. According to Bugbee [1996], scores from conventional and computer based testings can be considered as equivalent when the order of the scores is approximately the same and the scores are, or have been made approximately the same by scale adjustment. Hypothesis B.1 states hence that the score ranking for the four tested perceptual modalities is the same for GETOLS and for conventional testing.
Still, we think that the results achieved via GETOLS might as well be similar concerning scores (Hypothesis B.2a) or be even higher concerning scores (Hypothesis B.2b).

Hypothesis C states that testing via GETOLS is more time efficient than the conventional testing.

5.5 The GETOLS Test Setting

We decided to do our testing with 12-13 years old pupils. At that age, adolescents show high interest and engagement when interacting with computers [KIM-Studie 2010], and are old enough to provide useful feedback on the game and the test design. To ensure that all test persons are experienced in the handling of computers, we chose a school type with computer science classes. From a very early stage in game development on, we chose a specific teacher as contact teacher, which allowed us to get feedback from a field expert.

5.5.1 The Didactic Adventure Game "Save the city"

Following a suggestion of our contact teacher, we decided to focus on environmental protection as a subtopic of the subject matters "city and ecology", which is taught in biology at 7th grade. The game is about a young boy who is told that he had been chosen to save the city. In order to do so, he needs to complete several missions.

We chose the genre of adventure games as a test setting. Adventure games are intuitive and easy to play. According to Rapeepisarn et al. [2008], adventure games are suitable for discovery learning and guided learning, role playing, coaching and intelligent tutors. Their focus on the storyline helps to build up suspense and game engagement and to keep up the player's concentration.

The contents are prepared according to Vester [2011]. The story is made as suspenseful and fascinating as possible so as to enhance learning motivation. We pay attention to address all learning types during the game by not only using visual and kinesthetic stimulation (which are part of adventure games as they usually have a graphical interface with an appealing story environment design where the main character needs to be moved around during the game and to interact with in-game characters and objects), but also by addressing the auditory system via complete speech recording and providing textual input via subtitles.

The game can be played in about half an hour, which is short enough to be performed during one school lesson. The gameplay is designed to evoke a flow feeling that permits to keep up concentration for the duration of the testing as described by Kiili [2005b].

FIGURE 28 SCREENSHOTS OF THE EDUCATIONAL ADVENTURE GAME "SAVE THE CITY"

5.5.2 Integrating the Testing Method into the Game

Therefore it is necessary to not only loosely combine the test and the game, but to really embed the testing sequences into the game. The testing sequences as described in [Vester 2011] always follow the same test procedures:

- Memorization phase: 10 objects are presented according to the perceptual modality that shall be tested
- Different brain activity: mental arithmetic
- Test phase: test the memorized objects

Due to organizational issues, it is necessary to define a surrounding phase:

- Relaxation: wait until everything is prepared for the next testing phase; this part applies before and after each testing loop

Those patterns are mapped to the game as follows:

The story itself shall enhance engagement and help to focus on the game; furthermore it shall allow relaxation between the testing phases. In order to raise the level of attention, every memorization phase is preceded by an unexpected event in the story.

During the memorization phase, 10 objects are presented according to the perceptual modality that shall be tested: via pictures (visual), dictation (auditory), videos of use (kinesthetic) and as words (textual). Each object is shown for a period of 2 seconds.

The phase of different brain activity uses an adapted concept. According to Vester [2011], the goal of this phase is productive distraction. Instead of 30 seconds of mental arithmetic, we implemented four different minigames, each lasting 30 seconds. The tasks of the minigames are to identify differences between two pictures, to reassemble the pieces of a diamond, to find the way through a labyrinth and to click several times at a moving object. While mental arithmetic might resemble an examination situation and hence cause stress, our minigames are designed to be a fun activity of distraction.

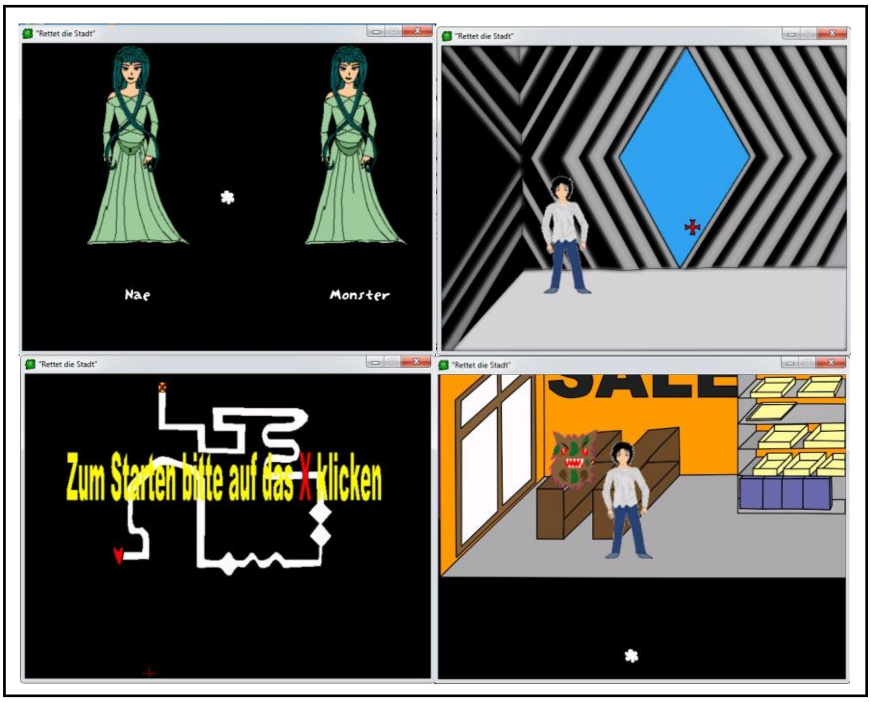

FIGURE 29 SCREENSHOTS OF THE FOUR GETOLS MINIGAMES.

Furthermore, the minigames are part of the gameplay and add another element to the storyline. Figure 29 shows screenshots of the minigames.

In the testing phase, the player is asked to recall as many of the 10 objects presented in the memorization phase as possible and to list them by keyboard input.

The testing procedure can hence be depicted as shown in Figure 30.

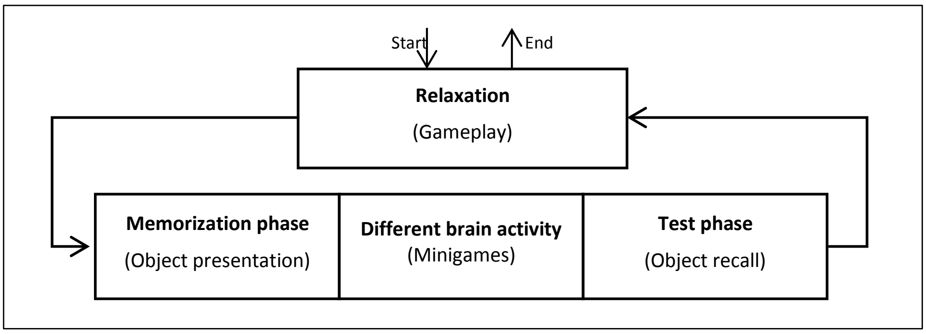

FIGURE 30 THE GETOLS TESTING PROCEDURE

5.5.3 Implementation and Testing

"Save the city" was implemented in SLUDGE, an object oriented programming language designed for the programming of adventure games.[24] Animated graphics were produced with the help of Anime Studio Pro.[25]

The test results for each test phase are stored as a list of words and are, at this stage of development, evaluated manually. An automated evaluation can be realized by integration of spell checking and synonym treatment, but might still need manual verification in order to assure correct evaluation of results.

The alpha testing phase was started after implementation of basic functionality. Goals of this phase were to collect feedback, to eliminate logical flaws, and to find technical as well as usability issues. We followed the "think-aloud" approach, as described by Young [2005], which is proven to be an accurate data-collection method [Cooke 2010]. 28 people of different age, with different expertise in the manipulation of computers and different experience of playing computer games, were asked to play and comment the game. The so obtained feedback led to more explicit information about the time limits during the testing sequence. Furthermore, we reduced the resolution of the videos in order to ensure functionality also on less performant computers.

The first loop of beta testing was already conducted with our final test subjects. Each subject was asked to play a very short sequence from the game for testing. In order to avoid measurement bias, the memorization phases were executed with different objects than in the final version. As several test subjects had problems with the use of

[24] http://opensludge.sourceforge.net/
[25] http://anime.smithmicro.com/

the inventory during the last minigame, we redesigned the minigame. The second loop of beta testing was conducted under the same conditions as the first one.

In a last beta testing phase, 9 computer scientists tested the game with regard to functionality and usability.

5.6 Experimental Results

We opted for repeated measures design. In our case, this means the application of different test methods to the same test persons and hence the elicitation of different result samples that can be compared to each other. The method comparison leads to a proof of convergent validity, which can be referred to as similarity of results gained with different test methods. Construct validation is based on the analysis of differences and similarities of the test results per test subject, elicited with different test methods. The lower variance allows to work with fewer test subjects [Mellinger and Hanson 2017]. One set of data was recorded via GETOLS method, the second set was collected in a conventional test setting, where a psychologist conducted the learning strategy test. We had 24 test subjects, whereof 18 test subjects were recorded via GETOLS and 16 tested by a psychologist. The conventional test was conducted one week before the GETOLS screening. We chose within-subjects analysis for the comparison of results. As only 10 test persons were recorded in both tests, we base our analysis upon the results of those 10 test subjects.

All test subjects were 12 or 13 years of age.

The following factors were considered when setting up the test:

- Location: The location chosen for the testing were the computer labs of the school. In addition to organizational advantage, the place is known to the subjects, which reduces fear, and can be closed for the public in order to avoid distraction.
- Preparation: Test preparation is finished before the test subject enters the room.
- Test procedure: Each test subject is welcomed individually in order to reduce tension.
- Duration: Even though average testing only takes 45 minutes, testing is scheduled for two hours per person in order to avoid stress.

The results of the testing are shown in Table 16-18.

TABLE 16 COMPARISON OF METHODS: RESULTS BY ORDER

<table>
<tr><th></th><th>Method</th><th colspan="4">Perceptual Modality</th><th colspan="3">Similarity: Subjects</th></tr>
<tr><th></th><th>g = GETOLS
p = Psychologist</th><th>Text based</th><th>Auditory</th><th>Visual</th><th>Kinesthetic</th><th><g,p></th><th>||g|| * ||p||</th><th>cos(g,p)</th></tr>
<tr><td rowspan="2">Subject 1</td><td>g</td><td>3</td><td>4</td><td>1</td><td>2</td><td rowspan="2">27</td><td rowspan="2">30.000</td><td rowspan="2">0.900</td></tr>
<tr><td>p</td><td>4</td><td>2</td><td>1</td><td>3</td></tr>
<tr><td rowspan="2">Subject 2</td><td>g</td><td>3</td><td>4</td><td>1</td><td>2</td><td rowspan="2">24</td><td rowspan="2">29.749</td><td rowspan="2">0.807</td></tr>
<tr><td>p</td><td>3.5</td><td>2</td><td>3.5</td><td>1</td></tr>
<tr><td rowspan="2">Subject 3</td><td>g</td><td>3</td><td>4</td><td>2</td><td>1</td><td rowspan="2">29.5</td><td rowspan="2">29.749</td><td rowspan="2">0.992</td></tr>
<tr><td>p</td><td>2.5</td><td>4</td><td>2.5</td><td>1</td></tr>
<tr><td rowspan="2">Subject 4</td><td>g</td><td>1</td><td>4</td><td>1</td><td>3</td><td rowspan="2">28</td><td rowspan="2">28.222</td><td rowspan="2">0.992</td></tr>
<tr><td>p</td><td>1.5</td><td>4</td><td>1.5</td><td>3</td></tr>
<tr><td rowspan="2">Subject 5</td><td>g</td><td>1.5</td><td>4</td><td>1.5</td><td>3</td><td rowspan="2">25.5</td><td rowspan="2">29.749</td><td rowspan="2">0.857</td></tr>
<tr><td>p</td><td>4</td><td>3</td><td>1</td><td>2</td></tr>
<tr><td rowspan="2">Subject 6</td><td>g</td><td>2.5</td><td>4</td><td>2.5</td><td>1</td><td rowspan="2">29.5</td><td rowspan="2">29.500</td><td rowspan="2">1.000</td></tr>
<tr><td>p</td><td>2.5</td><td>4</td><td>2.5</td><td>1</td></tr>
<tr><td rowspan="2">Subject 7</td><td>g</td><td>3</td><td>4</td><td>2</td><td>1</td><td rowspan="2">28</td><td rowspan="2">28.983</td><td rowspan="2">0.966</td></tr>
<tr><td>p</td><td>2</td><td>4</td><td>2</td><td>2</td></tr>
<tr><td rowspan="2">Subject 8</td><td>g</td><td>3</td><td>4</td><td>1</td><td>2</td><td rowspan="2">30</td><td rowspan="2">30.000</td><td rowspan="2">1.000</td></tr>
<tr><td>p</td><td>3</td><td>4</td><td>1</td><td>2</td></tr>
<tr><td rowspan="2">Subject 9</td><td>g</td><td>4</td><td>3</td><td>1.5</td><td>1.5</td><td rowspan="2">24</td><td rowspan="2">28.740</td><td rowspan="2">0.835</td></tr>
<tr><td>p</td><td>3</td><td>1</td><td>3</td><td>3</td></tr>
<tr><td rowspan="2">Subject 10</td><td>g</td><td>3</td><td>4</td><td>2</td><td>1</td><td rowspan="2">29.5</td><td rowspan="2">29.749</td><td rowspan="2">0.992</td></tr>
<tr><td>p</td><td>2.5</td><td>4</td><td>2.5</td><td>1</td></tr>
<tr><td rowspan="3">Similarity:
Perceptual Modality</td><td><g,p></td><td>76.25</td><td>119</td><td>32.75</td><td>37.5</td><td></td><td></td><td></td></tr>
<tr><td>||g|| * ||p||</td><td>83.285</td><td>132.068</td><td>36.297</td><td>39.481</td><td></td><td></td><td></td></tr>
<tr><td>cos(g,p)</td><td>0.916</td><td>0.901</td><td>0.902</td><td>0.950</td><td></td><td></td><td></td></tr>
</table>

TABLE 17 COMPARISON OF METHODS: RESULTS BY SCORES

<table>
<tr><th></th><th>Method</th><th colspan="4">Perceptual Modality</th><th colspan="3">Similarity: Subjects</th></tr>
<tr><th></th><th>g = GETOLS
p = Psychologist</th><th>Text based</th><th>Auditory</th><th>Visual</th><th>Kinesthetic</th><th><g,p></th><th>||g|| * ||p||</th><th>cos(g,p)</th></tr>
<tr><td rowspan="2">Subject 1</td><td>g</td><td>4</td><td>3</td><td>9</td><td>6</td><td rowspan="2">148</td><td rowspan="2">155.82683</td><td rowspan="2">0.94977</td></tr>
<tr><td>p</td><td>4</td><td>7</td><td>9</td><td>5</td></tr>
<tr><td rowspan="2">Subject 2</td><td>g</td><td>5</td><td>3</td><td>7</td><td>6</td><td rowspan="2">99</td><td rowspan="2">105.19981</td><td rowspan="2">0.94107</td></tr>
<tr><td>p</td><td>4</td><td>5</td><td>4</td><td>6</td></tr>
<tr><td rowspan="2">Subject 3</td><td>g</td><td>4</td><td>2</td><td>5</td><td>6</td><td rowspan="2">87</td><td rowspan="2">87.72115</td><td rowspan="2">0.99178</td></tr>
<tr><td>p</td><td>5</td><td>3</td><td>5</td><td>6</td></tr>
<tr><td rowspan="2">Subject 4</td><td>g</td><td>8</td><td>6</td><td>8</td><td>7</td><td rowspan="2">207</td><td rowspan="2">207.42710</td><td rowspan="2">0.99794</td></tr>
<tr><td>p</td><td>8</td><td>5</td><td>8</td><td>7</td></tr>
<tr><td rowspan="2">Subject 5</td><td>g</td><td>7</td><td>2</td><td>7</td><td>4</td><td rowspan="2">104</td><td rowspan="2">118.49895</td><td rowspan="2">0.87764</td></tr>
<tr><td>p</td><td>3</td><td>5</td><td>7</td><td>6</td></tr>
<tr><td rowspan="2">Subject 6</td><td>g</td><td>5</td><td>4</td><td>5</td><td>7</td><td rowspan="2">111</td><td rowspan="2">111.44505</td><td rowspan="2">0.99601</td></tr>
<tr><td>p</td><td>5</td><td>3</td><td>5</td><td>7</td></tr>
<tr><td rowspan="2">Subject 7</td><td>g</td><td>5</td><td>3</td><td>7</td><td>8</td><td rowspan="2">109</td><td rowspan="2">111.12156</td><td rowspan="2">0.98091</td></tr>
<tr><td>p</td><td>5</td><td>3</td><td>5</td><td>5</td></tr>
<tr><td rowspan="2">Subject 8</td><td>g</td><td>5</td><td>4</td><td>8</td><td>6</td><td rowspan="2">110</td><td rowspan="2">110.11812</td><td rowspan="2">0.99893</td></tr>
<tr><td>p</td><td>4</td><td>3</td><td>6</td><td>5</td></tr>
<tr><td rowspan="2">Subject 9</td><td>g</td><td>3</td><td>6</td><td>8</td><td>8</td><td rowspan="2">131</td><td rowspan="2">138.57489</td><td rowspan="2">0.94534</td></tr>
<tr><td>p</td><td>5</td><td>6</td><td>5</td><td>5</td></tr>
<tr><td rowspan="2">Subject 10</td><td>g</td><td>4</td><td>2</td><td>6</td><td>7</td><td rowspan="2">112</td><td rowspan="2">113.64418</td><td rowspan="2">0.98553</td></tr>
<tr><td>p</td><td>5</td><td>3</td><td>5</td><td>8</td></tr>
<tr><td rowspan="3">Similarity:
Perceptual Modality</td><td><g,p></td><td>235</td><td>156</td><td>410</td><td>366</td><td></td><td></td><td></td></tr>
<tr><td>||g|| * ||p||</td><td>257.721</td><td>171.216</td><td>433.274</td><td>401.186</td><td></td><td></td><td></td></tr>
<tr><td>cos(g,p)</td><td>0.912</td><td>0.911</td><td>0.946</td><td>0.912</td><td></td><td></td><td></td></tr>
</table>

Table 18 Comparison of Methods: Score Differences

	Method	Perceptual Modality					Differences: Subject			
	g = GETOLS p = Psychologist	***Text based***	***Auditory***	***Visual***	***Kinesthetic***	⦰	***T***	***A***	***V***	***K***
Subject 1	*g*	4	3	9	6	5.5	**0**	-4	**0**	**1**
	p	4	7	9	5	6.25				
Subject 2	*g*	5	3	7	6	5.25	**1**	-2	**3**	**0**
	p	4	5	4	6	4.75				
Subject 3	*g*	4	2	5	6	4.25	-1	-1	**0**	**0**
	p	5	3	5	6	4.75				
Subject 4	*g*	8	6	8	7	7.25	**0**	**1**	**0**	**0**
	p	8	5	8	7	7				
Subject 5	*g*	7	2	7	4	5	**4**	-3	**0**	-2
	p	3	5	7	6	5.25				
Subject 6	*g*	5	4	5	7	5.25	**0**	**1**	**0**	**0**
	p	5	3	5	7	5				
Subject 7	*g*	5	3	7	8	5.75	**0**	**0**	**2**	**3**
	p	5	3	5	5	4.5				
Subject 8	*g*	5	4	8	6	5.75	**1**	**1**	**2**	**1**
	p	4	3	6	5	4.5				
Subject 9	*g*	3	6	8	8	6.25	-2	**0**	**3**	**3**
	p	5	6	5	5	5.25				
Subject 10	*g*	4	2	6	7	4.75	-1	-1	**1**	-1
	p	5	3	5	8	5.25				
Average ⦰	***g***	5	3.5	7	6.5	5.5	0.2	-0.8	1.1	0.5
	p	4.8	4.3	5.9	6	5.25				

Looking at the test results, our hypotheses lead to the following conclusions:

The GETOLS test method can be considered equal to the conventional test setting. As can be seen in Table 16, the order of the scores is similar for the two test methods. Similarity was calculated as cosine similarity as described in Section III.B. All results show a cosine similarity greater than 0.8, and 70% have a similarity greater than 0.9. Hypothesis B.1 can hence be verified. Concerning the scores, the similarity of results based on scores is even higher than based on order. The lowest vector similarity of scores is 0.88, and 90% have a similarity greater than 0.9 (see Table 17). Hypothesis B.2a can hence be verified as well. We could yet not prove hypothesis B.2b, which states that test results via GETOLS are higher; this is only the case for 75% of the results (see Table 18), which is not high enough for significant conclusions. A slight scale adjustment might hence be taken into consideration when using the GETOLS method. However, in Vester's learning strategies test, the perceptual modalities' order is more informative than the absolute scores [Vester 2011], which makes hypothesis B.1 the most important. These results also support Hypothesis A: Vester's learning strategies test can thus be performed digitally via the GETOLS method.

Furthermore, the GETOLS test method is more time efficient than the conventional testing approach: While the conventional testing needs a human test supervisor and can only be executed in a serial manner, the GETOLS method allows parallel testing, which shows time efficiency effects starting from a test group of more than two test takers. In our experiment, testing 16 persons with the conventional testing method took 130 minutes, whereas testing via the game "Save the city" took 45 minutes for 18 test persons and would not take longer even for a bigger group (of course, the group size is limited according to the number of computers available).

As suggested by Vester [2011], the phase of different brain activity was accomplished with mental arithmetic exercises in the conventional test setting. According to the teacher, that phase was experienced by most of the test subjects as an examination situation and hence as quite stressful. This might be a reason for the lower test scores in the conventional test setting. Oral feedback by the test subjects confirms that assumption, which corresponds to the findings of Sarason [1984]. Furthermore, several test subjects pointed out that playing the minigames was much more fun. It would be interesting to systematically test those aspects in a second experimental run.

CHAPTER 6

Discussion and Conclusion

There are things known and there are things unknown,
and in between are the doors of perception.
(Aldous Huxley)

The last chapters showed that perceptual preference can be elicited in manifold ways. This chapter brings the different methods together and looks for structured answers to our research questions. Moreover, we discuss our results, the chosen research methods and their limitations as well as opportunities for further research.

6.1 Answers to our Research Questions

Along with the development and evaluation of the different methods for perceptual preference testing came a number of research questions, raised in the last three chapters. This section answers all these questions and can thus be seen as an unconventionally structured attempt to a summary.

According to the findings of our studies, the questions we raised in the introduction can be answered as follows:

How can perceptual preference be tested? This thesis provides three new methods to elicit perceptual preference, the Perceptual Preferences Questionnaire (PPQ), the analysis of word use as well as the Game Embedded Testing of Learning Strategies

(GETOLS). Each approach is validated with statistical methods as outlined in the previous chapters.

What interrelations exist between perceptual preference and interests of a user? Our results show that there are several significant correlations between perceptual preference and interests of a user. Looking at the strong correlation between forums and the interest most close to the forum topic, one can assume that active membership to a topic related forum can be treated as expression of interest in that topic. Furthermore, we found a strong correlation between membership in a music forum and auditory preference. Combining perceptual preference and interests, there is only one positive correlation between music and auditory preference, furthermore negative correlations concerning interests in music, art, psychology, nature, and travelling. Due to the rather weak reliability of the auditory scales, the results still should be interpreted with caution. Therefore, our results cannot yet provide a significant proof of the interrelation between interests and perceptual preference, but rather offer a first insight. Although positive correlations allow more concrete implications in the form of *if…then*, negative correlations can be interpreted as guidelines in the form of *if…then don't*, which still helps to reduce options. Given a user whose interests are, e.g., music, nature, and travelling, one can assume that this user has a high auditory preference (positive correlation between music and auditory preference) and a low visual preference (negative correlation between music and nature and visual preference) and a low kinesthetic preference (negative correlation between nature and travelling and kinesthetic preference), whereas no information is provided by those three interests about visual and olfactory/gustatory preference. This user could hence be considered as to be reached best via auditory information. With the help of rule-based or machine learning inference methods (which are planned as future work), one could conclude perceptual preference patterns even without complete knowledge of the user's interests. However, for that purpose, the findings of this thesis need to be substantiated with a larger dataset.

Is sensory vocabulary used sufficiently often to be considered as an indicator for individual preference? Looking at the distribution of sensory expression within the E30 forum corpus, around one out of 30 terms could be classified as sensory expression. (By manual analysis we found that there is still a large number of non-systematic "dummy terms" within the corpus replacing original links, images, references to other postings etc. The occurrence of sensory expression might thus be higher in text corpora containing only "real" words). Even though the used number of sensory terms might not seem very high, it is sufficient to obtain relevant information on the average user (who publishes 35 posts). As such, it is comparable to the measures behind the software LIWC (linguistic inquiry and word count) and its use for computer supported automatic detection of personality features [Boyd and Pennebaker 2017].

Does the use of sensory vocabulary stay consistent within the posts of one user? Our results show that authors of forum posts have individual tendencies concerning the use of sensory expression, using sensory expression patterns in similar distributions over time. Text can thus be handled as a source of implicit information about its author through word by word analysis and identification of all sensory expressions. Hence, that distribution can be considered as an interesting extension to user descriptions for user modeling. The corpus analysis further showed a more similar distribution of sensory vocabulary within answer threads than in randomly chosen pairs of posts. This might be explained either by a tendency of users to adapt their use of expression to the dialog partner or as an effect of topic relatedness. In any case, there is a tendency for authors to rather answer posts that correspond to their personal sensory preference.

How can the use of sensory vocabulary (and analogously perceptual preference) be modeled? We suggest the structure of a vector with four dimensions for each user. The vector describes the outcome of either a self-report elicited via our questionnaire, or how much a user expresses his or her thoughts in visual, auditory, kinesthetic and olfactory and gustatory terms, or the user's preferred learning style concerning object memorization respectively.

Is it possible to embed a psychological test into a game? As the results obtained by a psychologist and the ones calculated by our GETOLS method show a high similarity, the answer is yes: it actually is possible to embed a psychological test into a game. Results are similar in terms of scores and score order for both testing methods.

How could a psychological test be adjusted in order to make it feel game-like? We chose the testing method of Vester after analyzing several learning styles testing methods by the criteria of technical applicability, game integratability, complexity of test evaluation, usefulness of results, and name recognition.

What are the benefits of game embedded testing? The benefits of game embedded testing are certainly a more time efficient testing. Furthermore, the feedback of our test subjects indicates a higher fun factor and less stress. Those effects should be elicited directly in a second run.

How can such knowledge be used? We suggest to extend user models for a variable on perceptual preference. The rest of this answer bridges this section with the research outlook and is hence continued in section 6.3.

6.2 Limitations

Looking at the PPQ, our experiment with 76 participants provides a sufficient sample size to allow psycho-pedagogical sound interpretation of data. Our sample varies quite well concerning age, but in terms of gender male participants are slightly over-represented. A repetition of our study with a higher number of participants might give further insights, and might aim towards a scale revision of visual and auditory scales. As, in this thesis, data analyses were based on the same sample to validate the questionnaire and to calculate the respective correlations, such a repeated study should be based on two different sample groups in order to control for possible confounding side effects. An unequal response considering the different types of forums (an overrepresentation of music forum members) might have slightly effected the correlations of the results. While for our research focus and in order to assure the representation of all four preferences types it was necessary to only admit participants with a certain forum background, it would be interesting to analyze a sample with participants with different backgrounds, either by including more forums or by also contacting users who are not active in any internet forum.

In regard to vocabulary of perceptual preference, our corpus of sensory expression is still far from covering all expressions with a direct link to a sensory system that exist in German. Therefore, fewer words are classified as sensory expression than there are effectively within the analyzed texts. The method of matching letters around an identified sensory stem with components used in German word formation helps to classify also words that are not spelled correctly and ad hoc word creations. On the other hand, it might sometimes allow compositions that should not be classified as sensory expression. In every such case we found, we implemented exception handling. Nevertheless it is possible that due to this, some more words are classified as sensory expression than are effectively within the analyzed texts. However, all these misclassifications are constant during the whole corpus analysis and should therefore not affect any of the presented results. At the moment, we only examined the use of sensory vocabulary within rather narrow discussion topics. Further research needs to take different topics into account, in order to investigate the influence of topic on the use of sensory expression.

As concerns our GETOLS methodology, our GETOLS method has only been realized with one testing method, namely Vester's learning strategies test. Nevertheless, other test methods could be implemented by the same principle, most likely accompanied by the need of adjusting the storyline of the game to the testing method. On the infrastructural side, a computer is needed for conducting the test. However, computers are easily available and are thus a good substitute for the conventional testing material that needs to be assembled and prepared first in order to conduct a test in the conventional way.

So far, our studies with direct user interaction were conducted with a rather low number of test persons. Therefore, especially the GETOLS results (n=10) cannot yet provide a significant proof of our method, but rather offer a first insight. Therefore, it would be of great interest to repeat our studies with a higher number of test persons, including people with more diverse backgrounds and of different age.

In this thesis, we discuss first implications on how to present information to a specific user in order to ease understanding, which can be achieved by matching his or her perceptual preferences. Of course, such suggestions need further grounding. Our next steps in this regard are thus to identify blocks of content suitable for adaptation to perceptual preference and to design different versions in order to test those adaptions with users.

6.3 Cross Connections and Outlook

Following the classification structure on the design of user models as suggested by Brun, Boyer and Razmerita [2010], the investigation in perceptual preference as a new variable for user modeling can be seen as an extension to existing user models.

Such knowledge about the users' perceptual preferences might be of interest for every user model used in a setting where the user's interest needs to be captured and/or the user's process of information perception and organization shall be supported. It can be used for adapting presentation of information to the user's perceptual preference, which has been proven to support understanding and learning [Dunn 1988; Tsianos et al. 2008]. This can be done by choosing different forms of content presentation (like visual versus textual content) [Koć-Januchta et al. 2017], or, taking the findings of psycholinguistic research into account, also by reformulating verbal information in regard to the perceptual preference of a user by using perceptually stimulating vocabulary. It might hence be an interesting extension for adaptive hypermedia, especially for the field of flexible content and interface design. Such personalization could be useful for a wide range of applications including, but not limited to, e-commerce and e-learning. However, the degree of adaption should be chosen according to the specific goals of an application: if an application only aims to attract the user's attention, one should apply a higher degree of adaptation to the user's perceptual preference, than in an educational tool that does not only seek to ease understanding as a short-term effect, but also tries to reach long-term effects by training the user's less preferred perceptual systems instead of presenting each content in the same way and thus creating "perceptual filter bubbles".

We furthermore suggest a modeling technique using different ways of data acquisition, namely explicit data acquisition via the Perceptual Preference Questionnaire (PPQ), implicit data acquisition via analysis of the use of sensory vocabulary in forums, and embedded data acquisition via the Game Embedded Testing of Learning Strategies (GETOLS) method.

The obtained results of our studies are quite encouraging. Next steps could thus be to repeat all studies with a larger number of users in order to ground our first insights on larger datasets. That would allow deeper understanding of the relation between interests and perceptual preference and would provide data to revise the PPQ in order to assure high reliability of scales. Another interesting extension would be to enlarge the corpus of sensory expression, to systematically investigate the relation between preferred sensory system(s) and the use of sensory expression by combining forum text analysis with user tests on perceptual preference, and examine the influence of topic on the use of sensory expression. For game embedded testing, it could be beneficial to support automated test evaluation, to monitor reactions of stress and engagement for both testing methods, and to investigate the relation of test results and in-game behavior.

In addition, it is envisioned to investigate further implicit preference elicitation on a textual basis as an additional access key to perceptual preference, to investigate methods to deduce perceptual preference patterns from a known interest profile, to examine interrelations between perceptual preference and other user-modeling related factors, and to design and test content adaptation to perceptual preferences.

Even though direct measurement of sensation is quite challenging [Laming 1997], it might be interesting to compare the results elicited with our testing methods to experimental testing of information processing.

It would be most gratifying to the author if this thesis might have awakened or encouraged the desire to further investigate the topic of perceptual preference and its application, addressing some of the above-mentioned aspects or others chosen from the wide range of opportunities in this field.

6.4 Summary

In this theses, we developed three different methods for testing perceptual preference.

The first method is an explicit testing method. Based on findings from cognitive information processing and learning styles, we developed the Perceptual Preference Questionnaire (PPQ), which was validated with statistical methods based on the results of an online survey with 76 participants. We examined co-occurrence patterns of perceptual preference and interests. An analysis of the so gained data revealed significant correlations between perceptual preference and interests. As a result, a given list of interests of a user can be treated as an indication to his or her perceptual preference. A user’s interest in, e.g., music can thus be interpreted as a high indication to auditory preference. Furthermore, we identified several correlations between a user’s interests and membership in certain forums. These findings support our approach of examining perceptual preference and interests jointly.

The second part explores user generated text as an implicit source of information. We proposed a new dimension for user modeling based on the use of sensory expressions. Based on findings from cognitive information processing and learning styles, we investigated the potential of the idea to analyze the use of sensory expression as an individual preference that might indicate perceptual preference. We opted for an implicit approach to data acquisition concerning the use of sensory vocabulary by means of analyzing forum text. Starting from a list of 356 stems of German sensory expressions, we implemented a testing pipeline in order to classify more than 1,000,000 forum posts with regard to the use of sensory vocabulary. We identified more than 20,000 different sensory terms (by counting difference in at least one letter as distinction criterion). We found that authors tend to use sensory expressions in similar distributions when writing new posts. Furthermore, similarity concerning the use of sensory vocabulary is higher within answer threads. Both of those results support our idea of treating the use of sensory vocabulary as an interesting new dimension in user modeling.

In the third part on embedded testing, we proposed to embed a well-known learning style test into a rather unconventional test setting by integrating it into an educational computer game. We opted for Vester's learning strategies test as the most suitable for such an embedded testing approach, and used repeated measures design by comparing the data obtained by a psychologist with the data obtained via our game embedded testing of learning strategies (GETOLS) method. Our empirical validation showed a high similarity of results obtained with the two different test methods.

Overall, perceptual preference can be handled as an interesting extension for user models. Even though there is still a lot to do for further investigation on that topic, this thesis can be considered as a productive attempt to allow a first insight into the concepts of perceptual preference testing and its application, opening up a multitude of further research opportunities…

References

Adams, E. (2014a). Fundamentals of adventure game design. Berkeley, New Riders.

Adams, E. (2014b). Fundamentals of game design. Berkeley, New Riders.

Aleksić, V. and M. Ivanović (2017). "A Literature Review of Empirical Research on the Effects of Digital Games on Learning Styles and Multiple Intelligences." Croatian Journal of Education 19(2): 511-562.

All, A., E. P. Nuñez Castellar and J. Van Looy (2016). "Assessing the effectiveness of digital game-based learning: Best practices." Computers & Education 92: 90-103.

Amory, A. (2003). Another Country: Virtual Learning Spaces. World Conference on Educational Multimedia, Hypermedia and Telecommunications 2003. Lassner, D. and C. McNaught. Honolulu, Hawaii, USA, AACE: 708-715.

Amory, A. (2007). "Game object model version II: a theoretical framework for educational game development." Educational Technology Research and Development 55(1): 51-77.

Anderson, J. R., J. Funke, K. N. Oettingen and G. Plata (2013). Kognitive Psychologie. Berlin, Springer.

Atman, N., M. Inceoğlu and B. Aslan (2009). Learning Styles Diagnosis Based on Learner Behaviors in Web Based Learning. Computational Science and Its Applications – ICCSA 2009. Gervasi, O., D. Taniar, B. Murgante, A. Laganà, Y. Mun and M. Gavrilova. Berlin, Springer: 900-909.

Balahur, A., R. Mihalcea and A. Montoyo (2014). "Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications." Computer Speech & Language 28(1): 1-6.

Balakrishnan, V., N. Humaidi and E. Lloyd-Yemoh (2016). "Improving document relevancy using integrated language modeling techniques." Malaysian Journal of Computer Science 29(1): 45-55.

Balog, K. and M. de Rijke (2007). How to Overcome Tiredness: Estimating Topic-Mood Associations. Proceedings of the Int. Conf. on Weblogs and Social Media (ICWSM07): 199-202.

Barkan, R., S. Ayal and D. Ariely (2016). "Revisiting constructed preferences: Extrapolating preferences from relevant reminders." Decision 3(4): 281-294.

Bayardo, R. J., Y. Ma and R. Srikant (2007). Scaling up all pairs similarity search. Proceedings of the 16th international conference on World Wide Web (WWW). Banff, Alberta, Canada, ACM: 131-140.

Bellotti, F., B. Kapralos, K. Lee, P. Moreno-Ger and R. Berta (2013). "Assessment in and of serious games: an overview." Advances in Human-Computer Interaction 2013: 1-11.

Benamara, F., C. Cesarano, A. Picariello, D. Reforgiato and V. S. Subrahmanian (2007). Sentiment Analysis: Adjectives and Adverbs are Better than Adjectives Alone. Proceedings of the International Conference on Weblogs and Social Media (ICWSM).

Bond, M. and R. Beale (2009). What makes a good game? Using reviews to inform design. Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology. Cambridge, British Computer Society: 418-422.

Bortz, J. and N. Döring (2006). Forschungsmethoden und Evaluation. Berlin, Springer.

Bourkoukou, O., E. El Bachari and M. El Adnani (2016). "A Personalized E-Learning Based on Recommender System." International Journal of Learning and Teaching 2(2): 99-103.

Boyd-Graber, J. and P. Resnik (2010). Holistic sentiment analysis across languages: multilingual supervised latent Dirichlet allocation. Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics.

Boyd, R. L. and J. W. Pennebaker (2017). "Language-based personality: a new approach to personality in a digital world." Current Opinion in Behavioral Sciences 18: 63-68.

Bozsahin, C. (2002). The Combinatory Morphemic Lexicon. Ankara, Middle East Technical University, Turkey. PhD Thesis.

Braschler, M. and B. Ripplinger (2004). "How Effective is Stemming and Decompounding for German Text Retrieval?" Information Retrieval 7: 291-316.

Brun, A., A. Boyer and L. Razmerita (2010). Compass to Locate the User Model I Need: Building the Bridge between Researchers and Practitioners in User Modeling. User Modeling, Adaptation, and Personalization. De Bra, P., A. Kobsa and D. Chin. Berlin, Springer: 303-314.

Buckenhüskes, H. J., K. v. Heusinger and T. Nawrocki (2005). Die Dimensionen des Geschmacks. Bedeutungsfelder für Geschmacksadjektive. Semantik der Sinne. Linke, A. and J. Nuessli. Zürich, Labor für Lebensmittelchemie und -technologie: 75-84.

Bugbee, A. C. (1996). "The Equivalence of Paper-and-Pencil and Computer-Based Testing." Journal of Research on Computing in Education 28(3): 282-299.

Bühner, M. (2011). Einführung in die Test- und Fragebogenkonstruktion. München, Pearson.

Bülow, E. (1970). Versuch einer semantischen Faktorenanalyse der verbalen Ausdrücke des Sehens. Bonn. PhD Thesis.

Burgers, C., A. Eden, M. D. van Engelenburg and S. Buningh (2015). "How feedback boosts motivation and play in a brain-training game." Computers in Human Behavior 48: 94-103.

Calvert, G., C. Spence and B. E. Stein (2004). The handbook of multisensory processes. Massachusetts, MIT Press.

Canoo Engineering AG (2011). "Deutsche Wörterbücher und Grammatik." Retrieved 10.02.2012, from http://www.canoo.net/services/WordformationRules/ueberblick/.

Carstensen, K.-U., C. Ebert, C. Ebert, S. Jekat, H. Langer and R. Klabunde (2009). Computerlinguistik und Sprachtechnologie. München, Elsevier.

Chang, Y.-C., W.-Y. Kao, C.-P. Chu and C.-H. Chiu (2009). "A learning style classification mechanism for e-learning." Comput. Educ. 53(2): 273-285.

Chen, W. (2008). Dimensions of subjectivity in natural language. Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies. Columbus, Ohio, ACL: 13-16.

Chi, M., K. VanLehn, D. Litman and P. Jordan (2010). Inducing effective pedagogical strategies using learning context features. Proceedings of the 18th international conference on User Modeling, Adaptation, and Personalization. Big Island, HI, Springer-Verlag: 147-158.

Chung, J. and V. R. Rao (2012). "A General Consumer Preference Model for Experience Products: Application to Internet Recommendation Services." Journal of Marketing Research 49(3): 289-305.

Citron, F. M. M., C. Cacciari, M. Kucharski, L. Beck, M. Conrad and A. M. Jacobs (2016). "When emotions are expressed figuratively: Psycholinguistic and Affective Norms of 619 Idioms for German (PANIG)." Behavior Research Methods 48(1): 91-111.

Clark, D. B., E. E. Tanner-Smith and S. S. Killingsworth (2016). "Digital Games, Design, and Learning." Review of Educational Research 86(1): 79-122.

Clark, R. C. and R. E. Mayer (2016). E-learning and the science of instruction. Hoboken, Wiley.

Clément, D. (1971). Satzeinbettungen nach Verben der Sinneswahrnehmung im Deutschen. Probleme und Fortschritte der Transformationsgrammatik. Referate des 4. Linguistischen Kolloquiums Berlin, 6.-10. Oktober 1969. Wunderlich, D. München, Hueber: 245-265.

Clouet, E., R. Harastani, B. Daille and E. Morin (2015). "Compositional translation of single-word complex terms using multilingual splitting." Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 21(2): 263-291.

Coffield, F., D. Moseley, E. Hall and K. Ecclestone (2004). Learning styles and pedagogy in post-16 learning. A systematic and critical review. London LSRC.

Cohen, J. (1988). Statistical power analysis for the behavioural sciences. Hillsdale, Erlbaum.

Collins, R. J. (2011). Project Management. New York, Nova Science.

Cooke, L. (2010). "Assessing concurrent think-aloud protocol as a usability test method: A technical communication approach." IEEE Transactions on professional communication 53(3): 202-215.

Coren, S. (2003). Sensation and perception. Handbook of Psychology. Freedheim, D. K. New Jersey, Wiley: 85-108.

Crystal, D. (1987). The Cambridge Encyclopedia of Linguistics. New York, Oxford University Press.

Cuevas, J. (2015). "Is learning styles-based instruction effective? A comprehensive analysis of recent research on learning styles." Theory and Research in Education 13(3): 308-333.

Dashtipour, K., S. Poria, A. Hussain, E. Cambria, A. Y. A. Hawalah, A. Gelbukh and Q. Zhou (2016). "Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques." Cognitive Computation 8(4): 757-771.

Dastani, M., N. Jacobs, C. Jonker and J. Treur (2001). Modeling User Preferences and Mediating Agents in Electronic Commerce. Agent Mediated Electronic Commerce. Dignum, F. and C. Sierra. Berlin, Springer: 163-193.

Diersch, H. (1972). Verben der Fortbewegung in der deutschen Sprache der Gegenwart. Eine Untersuchung zu syntagmatischen und paradigmatischen Beziehungen des Wortinhalts. Berlin, Akademie-Verlag.

Dijkstra, T. and G. Kempen (1993). Einführung in die Psycholinguistik. Bern, Huber.

Dunn, R. (1988). "Commentary: Teaching Students Through Their Perceptual Strengths or Preferences." Journal of Reading 31(4): 304-309.

Egenfeldt-Nielsen, S. (2007). "Third Generation Educational Use of Computer Games." Journal of Educational Multimedia and Hypermedia 16(3): 263-281.

Engelberg, S. (2000). Verben, Ereignisse und das Lexikon. Tübingen, Niemeyer.

Erhel, S. and E. Jamet (2013). "Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness." Computers & Education 67: 156-167.

Ernst, M. O. and H. H. Bülthoff (2004). "Merging the senses into a robust percept." Trends in Cognitive Sciences 8(4): 162-169.

Escartín, C. P. (2014). Chasing the Perfect Splitter: A Comparison of Different Compound Splitting Tools. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC). Reykjavik, Iceland: 3340-3347.

Falkenberg, G. (1989). Einige Bemerkungen zu perzeptiven Verben. Wissen, Wahrnehmen, Glauben. Epistemische Ausdrücke und propositionale Einstellungen. Falkenberg, G. Tübingen, Niemeyer: 27-48.

Felder, R. M. (2002). "Learning and teaching styles in engineering education: Author's preface."

Felder, R. M. and L. Silverman (1988). "Learning and Teaching Styles In Engineering Education." Engr. Education 78(7): 674-668.

Feldman, J., A. Monteserin and A. Amandi (2014). "Detecting students' perception style by using games." Computers & Education 71: 14-22.

Ferreira Santos, E. N., G. Rodrigues Liska and M. Angelo Cirrillo (2017). "Methodology for constructing perceptual maps incorporating measuring error in sensory acceptance tests." Journal of Modern Applied Statistical Methods 16(1): 308-323.

Foerster, H. v. (1985). Das Konstruieren einer Wirklichkeit. Die erfundene Wirklichkeit. Was wissen wir, was wir zu wissen glauben? Beiträge zum Konstruktivismus. Watzlawick, P. München, Piper: 39-60.

Forcht, D. (2009). Systemtheorie und die Information der Systeme. Von Molekülen und Nervenzellen bis zur Gemeinschaft der Menschen. Hamburg, Diplomica.

Fricke, E. (1996). Die Verben des Riechens im Deutschen und Englischen. Eine kontrastive semantische Analyse. KIT-Report, TU Berlin.

Frings, S. and F. Müller (2013). Biologie der Sinne. Vom Molekül zur Wahrnehmung. Berlin, Springer.

Froschauer, J., M. Arends, D. Goldfarb and D. Merkl (2011). Towards an Online Multiplayer Serious Game Providing a Joyful Experience in Learning Art History. Third International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES11).

Galton, F. (1883). Human faculty and development. London, Macmillan.

Gasparetti, F. (2017). "Modeling user interests from web browsing activities." Data Mining and Knowledge Discovery 31(2): 502-547.

Gaussier, E. and F. Yvon, Eds. (2012). Textual information access: statistical models. Hoboken, Wiley.

Gawronski, B. and J. De Houwer (2014). Implicit measures in social and personality psychology. Handbook of research methods in social and personality psychology. Reis, H. T. and C. M. Judd. Cambridge, Cambridge University Press. 2: 283-310.

Gerling, M. and N. Orthen (1979). Deutsche Zustands- und Bewegungsverben. Eine Untersuchung zu ihrer semantischen Struktur und Valenz. Tübingen, Narr

Geyken, A. and T. Hanneforth (2006). TAGH: A Complete Morphology for German Based on Weighted Finite State Automata. Finite-State Methods and Natural Language Processing. Yli-Jyrä, A., L. Karttunen and J. Karhumäki. Berlin Springer: 55-66.

Goldberg, L. R. (2012). "Language and personality: Toward a taxonomy of trait descriptive terms." Psikoloji Çalışmaları Dergisi 12: 1-23.

Goldstein, E. B. and J. Brockmole (2016). Sensation & Perception. Boston, Cengage.

Golijani-Moghaddam, N., A. Hart and D. L. Dawson (2013). "The Implicit Relational Assessment Procedure: Emerging reliability and validity data." Journal of Contextual Behavioral Science 2(3): 105-119.

Gomaa, W. H. and A. A. Fahmy (2013). "A survey of text similarity approaches." International Journal of Computer Applications 68(13): 13-18.

Gregor, S. and A. R. Hevner (2013). "Positioning and presenting design science research for maximum impact." MIS quarterly 37(2): 337-355.

Gregory, R. L. and A. M. Colman (1995). Sensation and Perception. London Longman Essential Psychology.

Guntuku, S. C., D. B. Yaden, M. L. Kern, L. H. Ungar and J. C. Eichstaedt (2017). "Detecting depression and mental illness on social media: an integrative review." Current Opinion in Behavioral Sciences 18: 43-49.

Gupta, G. K. (2006). Introduction to Data Mining with Case Studies. New Delhi, Prentice-Hall of India.

Guy, I., U. Avraham, D. Carmel, S. Ur, M. Jacovi and I. Ronen (2013). Mining expertise and interests from social media. Proceedings of the 22nd international conference on World Wide Web. Rio de Janeiro, Brazil, ACM: 515-526.

Hamari, J., D. J. Shernoff, E. Rowe, B. Coller, J. Asbell-Clarke and T. Edwards (2016). "Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning." Computers in Human Behavior 54: 170-179.

Hamborg, F., N. Meuschke, A. Aizawa and B. Gipp (2017). Identification and Analysis of Media Biasin News Articles. Everything Changes, Everything Stays the Same? Understanding Information Spaces. Proceedings of the 15th International Symposium of Information Science (ISI 2017). Gäde, M., V. Trkulja and V. Petras. Berlin, Hülsbusch: 224-236.

Harley, T. A. (2017). Talking the talk. Language, Psychology and Science. Abingdon, Routledge.

Harm, V. (2000). Regularitäten des semantischen Wandels bei Wahrnehmungsverben des Deutschen. Stuttgart, Steiner.

Harras, G., E. Winkler, S. Erb and K. Proost (2004). Handbuch deutscher Kommunikationsverben. Teil 1: Wörterbuch. Berlin, De Gruyter.

Heagney, J. (2011). Fundamentals of Project Management. New York, Amacom.

Heckmann, H.-D. and F. Esken (1998). Generelle Einführung: Bewusstsein und Repräsentation. Bemerkungen über zwei Schlüsselbegriffe, ihre Ausdifferenzierung und ihren Zusammenhang. Bewusstsein und Repräsentation. Heckmann, H.-D. and F. Esken. Paderborn, Schöningh: 11-49.

Heinath, M., J. Dzaack, A. Wiesner and L. Urbas (2007). Applications for Cognitive User Modeling. User Modeling 2007. Conati, C., K. McCoy and G. Paliouras. Berlin / Heidelberg, Springer 127-136.

Herrmann, T. (1993). Mentale Repräsentation - ein erläuterungswürdiger Begriff. Mentale Repräsentation. Engelkamp, J. and T. Pechmann. Bern, Huber: 17-30.

Hess, W. (2002). "Grundlagen der Phonetik." Retrieved 28.05.2012, from http://www.ikp.uni-bonn.de/lehre/informationen-materialien/informationen-und-materialien-kopho/materialien-1/hess/grundlagen-der-phonetik.

Hevner, A. R. (2011). "Design Science Research in Information Systems." Retrieved 20.06.2012, from http://www.ec.tuwien.ac.at/files/u326/Vienna_2011_DSR_Presentation_hevner.pdf.

Hevner, A. R., S. T. March, J. Park and S. Ram (2004). "Design Science in Information Systems Research." MIS Quarterly 28(1): 75-105.

Hirumi, A., B. Appelman, L. Rieber and R. Van Eck (2010). "Preparing Instructional Designers for Game-Based Learning: Part 1." TechTrends 54(3): 27-37.

Holz, P. (2005). Die Sprache des Parfums. Eine empirische Untersuchung zur Grammatik, Metophorik und Poetizität des Parfumwerbetextes. Hamburg, Kovacz.

Horn, J. L. (1965). "A rationale and test for the number of factors in factor analysis." Psychometrika 30(2): 179-185.

Huang Ling, Y. (2011). Designing Serious Games to Enhance Political Efficacy and Critical Thinking Disposition for College Students: The Case of Taiwan. Games and Virtual Worlds for Serious Applications (VS-GAMES), 2011 Third International Conference on.

Huang, Y.-M., J.-P. Hwang and S. Y. Chen (2016). "Matching/mismatching in web-based learning: a perspective based on cognitive styles and physiological factors." Interactive Learning Environments 24(6): 1198-1214.

Hundsnurscher, F. (1976). Die Perzeptionsverben des Deutschen. Akten des V. Internationalen Germanisten-Kongresses, Cambridge 1975. Forster, L. and H.-G. Roloff. Bern, Lang: 66-73.

Hundsnurscher, F. (1977). Zur Rolle der Wortarten im System der Perzeptionswörter. Perspektiven der Wortbildungsforschung. Brekle, H. E. and D. Kastovsky. Bonn, Bouvier: 83-97.

Ingason, A., S. Helgadóttir, H. Loftsson and E. Rögnvaldsson (2008). A Mixed Method Lemmatization Algorithm Using a Hierarchy of Linguistic Identities (HOLI). Advances in Natural Language Processing. Nordström, B. and A. Ranta. Berlin, Springer 205-216.

Isbister, K. and N. Schaffer (2008). Game Usability: Advancing the Player Experience. Boca Raton, CRC Press.

Iten, N. and D. Petko (2016). "Learning with serious games: Is fun playing the game a predictor of learning success?" British Journal of Educational Technology 47(1): 151-163.

Iuppa, N. and T. Borst (2010). End-to-End Game Development: Creating Independent Serious Games and Simulations from Start to Finish. Amsterdam, Elsevier.

Jiang, B. and Y. Sha (2015). "Modeling Temporal Dynamics of User Interests in Online Social Networks." Procedia Computer Science 51: 503-512.

Johnson, R. B. and A. J. Onwuegbuzie (2004). "Mixed Methods Research: A Research Paradigm Whose Time Has Come." Educational Researcher 33(7): 14-26.

Jones, W. J. (2013). Historisches Lexikon deutscher Farbbezeichnungen. 5 Bände. Berlin, Akademie-Verl.

Joshi, A., P. Bhattacharyya and S. Ahire (2017). Sentiment Resources: Lexicons and Datasets. A Practical Guide to Sentiment Analysis. Cambria, E., D. Das, S. Bandyopadhyay and A. Feraco. Cham, Springer: 85-106.

Ju, E. and C. Wagner (1997). "Personal computer adventure games: their structure, principles, and applicability for training." SIGMIS Database 28(2): 78-92.

Jürgenson, I. B. (2009). Neuronale Korrelate phonotaktischer Verarbeitung. Berlin, Universitätsmedizin Berlin, Germany. PhD Thesis.

Kahneman, D. and A. Tversky (1982). "The psychology of preferences." Scientific American 246(1): 160-173.

Karampiperis, P., A. Koukourikos and G. Stoitsis (2014). Collaborative Filtering Recommendation of Educational Content in Social Environments Utilizing Sentiment Analysis Techniques. Recommender Systems for Technology Enhanced Learning: Research Trends and Applications. Manouselis, N., H. Drachsler, K. Verbert and O. C. Santos. New York, Springer: 3-23.

Kellermann, M. (1997). Suggestive Kommunikation. Unterschwellige Botschaften in Alltag und Werbung. Bern, Huber.

Kellner, G. (2010). Wege der Kommunikationsoptimierung. Anwendung von NLP im Bereich der Künstlichen Intelligenz. Saarbrücken, VDM.

Kellner, G. and B. Berendt (2011). Extracting knowledge about cognitive style. 7th International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE).

Kellner, G. and B. Berendt (2012). Towards a New Dimension for User Modeling: The Use of Sensory Vocabulary. Advances in User Modeling. Ardissono, L. and T. Kuflik. Berlin, Springer: 397-401.

Kellner, G. and M. Berthold (2012). I-Know my Users: User-Centric Profiling Based on Interests and Perceptual Preferences. Under Review.

Kellner, G. and J. Grünauer (2012). Algorithms for the Verification of the Semantic Relation Between a Word and a Given Word Fragment. i-know 2012. Graz, ACM Publishing.

Kellner, G. and A. Weißenbacher (2012). GETOLS: Game Embedded Testing of Learning Strategies. Digital Game and Intelligent Toy Enhanced Learning (DIGITEL), 2012 IEEE Fourth International Conference on.

Kemp, S. E., T. Hollowood and J. Hort (2009). Sensory evaluation. Oxford, Wiley-Blackwell.

Kern, M. L., J. C. Eichstaedt, H. A. Schwartz, L. Dziurzynski, L. H. Ungar, D. J. Stillwell, M. Kosinski, S. M. Ramones and M. E. P. Seligman (2013). "The Online Social Self." Assessment 21(2): 158-169.

Ketelhut, D. J., B. C. Nelson and C. Schifter (2009). SAVE SCIENCE. Situated assessment using virtual environments of science content and inquiry. AERA.

Khan, B. H. and M. Ally, Eds. (2015). International Handbook of E-Learning, Volume 1. Hoboken, Taylor and Francis.

Kiili, K. (2005a). Educational Game Design: Experiential gaming model revised. Research report. Tampere, Tampere University of Technology.

Kiili, K. (2005b). On educational game design. Building blocks of flow experience. Tampere, Tampere University of Technology. PhD Thesis.

Kiili, K., T. Lainema, S. de Freitas and S. Arnab (2014). "Flow framework for analyzing the quality of educational games." Entertainment Computing 5(4): 367-377.

KIM-Studie (2010). Basisstudie zum Medienumgang 6- bis 13-Jähriger in Deutschland, Medienpädagogischer Forschungsverband Südwest.

Klašnja-Milićević, A., M. Ivanović and A. Nanopoulos (2015). "Recommender systems in e-learning environments: a survey of the state-of-the-art and possible extensions." Artificial Intelligence Review 44(4): 571-604.

Koć-Januchta, M., T. Höffler, G.-B. Thoma, H. Prechtl and D. Leutner (2017). "Visualizers versus verbalizers: Effects of cognitive style on learning with texts and pictures – An eye-tracking study." Computers in Human Behavior 68(Supplement C): 170-179.

Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. New Jersey, Prentice-Hall.

Köper, M. and S. Schulte Im Walde Automatically Generated Affective Norms of Abstractness, Arousal, Imageability and Valence for 350 000 German Lemmas. Proceedings of the 10th International Conference on Language Resources and Evaluation, Portoroz, Slovenia.

Kornbrot, D. E. (2016). "Human Psychophysical Functions, an Update: Methods for Identifying their form; Estimating their Parameters; and Evaluating the Effects of Important Predictors." Psychometrika 81(1): 201-216.

Kozhevnikov, M., C. Evans and S. M. Kosslyn (2014). "Cognitive Style as Environmentally Sensitive Individual Differences in Cognition: A Modern Synthesis and Applications in Education, Business, and Management." Psychological Science in the Public Interest 15(1): 3-33.

Kruse, K. (2002). "The Benefits and Drawbacks of e-Learning." Retrieved 10.09.2011, from http://labsel.pesarosviluppo.it/docindexer/Uploads%5C213-Beginner%20Basics.doc.

Kurilovas, E., S. Kubilinskiene and V. Dagiene (2014). "Web 3.0–based personalisation of learning objects in virtual learning environments." Computers in Human Behavior 30: 654-662.

Laming, D. (1997). The Measurement of Sensation. New York, Oxford University Press.

Laserna, C. M., Y.-T. Seih and J. W. Pennebaker (2014). "Um . . . Who Like Says You Know." Journal of Language and Social Psychology 33(3): 328-338.

Lee, C. H., K. Kim, Y. S. Seo and C. K. Chung (2007). "The relations between personality and language use." The Journal of general psychology 134(4): 405-413.

Liang, S., Z. Ren, Y. Zhao, J. Ma, E. Yilmaz and M. D. Rijke (2017). "Inferring Dynamic User Interests in Streams of Short Texts for User Clustering." ACM Transactions on Information Systems (TOIS) 36(1): 1-37.

Liu, B. (2015). Sentiment analysis: mining opinions, sentiments, and emotions. Cambridge, Cambridge Univ. Press.

Liu, Y. and D. Ginther (1999). "Cognitive Styles and Distance Education." Online Journal of Distance Learning Administration 2(3).

Macherey, K., A. M. Dai, D. Talbot, A. C. Popat and F. Och (2011). Language-independent compound splitting with morphological operations. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, ACL: 1395-1404.

Martin, G. N. (2008). Psychology: A Beginner's Guide, Oneworld Publications.

Materynska, O. (2012). Typologie der Körperteilbenennungen. Frankfurt a.M., Peter Lang.

Medhat, W., A. Hassan and H. Korashy (2014). "Sentiment analysis algorithms and applications: A survey." Ain Shams Engineering Journal 5(4): 1093-1113.

Mellinger, C. D. and T. A. Hanson (2017). Quantitative research methods in translation and interpreting studies. London, Routledge.

Mercer, T. G., A. P. Kythreotis, Z. P. Robinson, T. Stolte, S. M. George and S. K. Haywood (2017). "The use of educational game design and play in higher education to influence sustainable behaviour." International Journal of Sustainability in Higher Education 18(3): 359-384.

Merriam-Webster (2003). Merriam-Webster's collegiate dictionary. Springfield, Merriam-Webster.

Moray, N. (2017). Attention. Selective processes in vision and hearing. London, Routledge.

Morris, P. E. and C. O. Fritz (2000). "The name game: Using retrieval practice to improve the learning of names." Journal of Experimental Psychology 6(2): 124-129.

Moser, R. B. (2000). A Methodology for the Design of Educational Computer Adventure Games. Computer Science and Engineering Sidney, University of New South Wales: 309.

Mou, Y. and W. Peng (2009). Gender and racial stereotypes in popular video games. Handbook of research on effective electronic gaming in education, IGI Global: 922-937.

Nawrocki, T. (2004). Wortsinn und Geschmackssinn – Geschmacksadjektive und Werbung. Stuttgart. Master Thesis.

Newell, E. E., S. K. McCoy, M. L. Newman, J. D. Wellman and S. K. Gardner (2017). "You Sound So Down: Capturing Depressed Affect Through Depressed Language." Journal of Language and Social Psychology: 1-24.

Nowson, S. (2006). The Language of Weblogs: A study of genre and individual differences. Edinburgh, University of Edinburgh. Doctoral Thesis.

O'Brien, G. and J. Opie (2004). "Notes toward a structuralist theory of mental representation." Representation in mind. New approaches to mental representation: 1-20.

Okan, Z. (2003). "Edutainment: is learning at risk?" British Journal of Educational Technology 34(3): 255-264.

Overlach, F. (2008). Sprache des Schmerzes – Sprechen über Schmerzen. Eine grammatisch-semantische und gesprächsanalytische Untersuchung von Schmerzausdrücken im Deutschen. Berlin, De Gruyter.

Oxford English Dictionary (2010). Oxford: Oxford University Press

Özpolat, E. and G. B. Akar (2009). "Automatic detection of learning styles for an e-learning system." Computers & Education 53(2): 355-367.

Paivio, A. (2008). Mental Representations. A Dual Coding Approach. Oxford, Oxford University Press.

Park, G., H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, M. Kosinski, D. J. Stillwell, L. H. Ungar and M. E. P. Seligman (2015). "Automatic personality assessment through social media language." Journal of Personality and Social Psychology 108(6): 934-952.

Parvez, S. and G. Blank (2008). Individualizing Tutoring with Learning Style Based Feedback. Intelligent Tutoring Systems. Woolf, B., E. Aïmeur, R. Nkambou and S. Lajoie. Berlin, Springer: 291-301.

Pask, G. (1976). "Styles and strategies of learning." The British Journal of Educational Psychology 46: 128 -148.

Patel, M. and A. Shah (2016). "An unsupervised stemming: A review." International Journal of Computer Science and Information Security 14(7): 476-489.

Payne, J. W., J. R. Bettman and D. A. Schkade (1999). "Measuring Constructed Preferences: Towards a Building Code." Journal of Risk and Uncertainty 19(1): 243-270.

Pivec, M. (2008). Game-based learning. Wien, Pabst Science Publ.

Plank, B. and D. Hovy (2015). Personality Traits on Twitter-or-How to Get 1,500 Personality Tests in a Week. Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2015), Lisboa, Portugal, ACL.

Popescu, E. (2010). Learning styles in technology enhanced education: latest trends and a case study. 5th International Conference on Virtual Learning (ICVL 2010).

Poria, S., E. Cambria, R. Bajpai and A. Hussain (2017). "A review of affective computing: From unimodal analysis to multimodal fusion." Information Fusion 37: 98-125.

Porter, M. F. (1980). "An algorithm for suffix stripping." Program 3: 130-137.

Porter, M. F., R. Boulton, P. Miles and e. al. (2002). "Snowball Project: German Stemming Algorithm." Retrieved 10.02.2012, from
http://snowball.tartarus.org/algorithms/german/stemmer.html.

Prensky, M. (2001). Digital game-based learning. New York, McGraw Hill.

Quinn, C. N. (2005). Engaging learning: Designing e-learning simulation games. San Francisco, CA, Pfeiffer.

Ramani, S., V. Sirigiri, N. L. Panigrahi and S. Sabharwal (2008). Games as Skins for Online Tests. Digital Games and Intelligent Toys Based Education, 2008 Second IEEE International Conference on.

Rapeepisarn, K., K. Wong, C. Fung and M. Khine (2008). The Relationship between Game Genres, Learning Techniques and Learning Styles in Educational Computer Games. Technologies for E-Learning and Digital Entertainment. Pan, Z., X. Zhang, A. El Rhalibi, W. Woo and Y. Li. Berlin, Springer: 497-508.

Reid, J. M. (1987). "The Learning Style Preferences of ESL Students." TESOL Quarterly 21(1): 87-111.

Ricci, F., L. Rokach and B. Shapira, Eds. (2015). Recommender Systems Handbook. New York, Springer.

Richling, J. (2008). Die Sprache in Foren und Newsgroups. Saarbrücken, VDM.

Robering, K. (1985). Die deutschen Verben des Sehens. Eine semantische Analyse. Göpping, Kümmerle.

Robotham, D. (1999). "The application of learning style theory in higher education teaching." Retrieved 10.09.2015, from
http://serprofessoruniversitario.pro.br/m%C3%B3dulos/c%C3%A9rebro-e-aprendizagem/application-learning-style-theory-higher-education-teaching-0.

Rodriguez, D. M., M. Teesson and N. C. Newton (2014). "A systematic review of computerised serious educational games about alcohol and other drugs for adolescents." Drug and Alcohol Review 33(2): 129-135.

Romero, M., M. Usart and M. Ott (2014). "Can Serious Games Contribute to Developing and Sustaining 21st Century Skills?" Games and Culture 10(2): 148-177.

Roos, E. (1975). Kollokationsmöglichkeiten der Verben des Sehvermögens im Deutschen und Englischen. Frankfurt a.M., Lang

Sadler-Smith, E. and R. Riding (1999). "Cognitive style and instructional preferences." Instructional Science 27: 355-371.

Salamon, G. (2017). Der Wortschatz der Wahrnehmung. Lexikologische Erhebung und semantisch-kategoriale Analyse. Vienna, University of Vienna. PhD thesis.

Santos, P. B. (2014). Using compound lists for german decompounding in a back-off scenario. Workshop on Computational, Cognitive, and Linguistic Approaches to the Analysis of Complex Words and Collocations (CCLCC 2014). Henrich, V. and E. Hinrichs. Tübingen, SfS: 51-55.

Sarason, I. G. (1984). "Stress, anxiety, and cognitive interference: reactions to tests." Journal of Personality and Social Psychology 46(4): 929-938.

Schell, J. (2015). The art of game design: A book of lenses. Burlington, Kaufmann.

Schepping, M.-T. (1982). Kontrastive semantische Analyse von Verben des Visuellen im Französischen und Deutschen. Tübingen, Narr

Schindling, E. (2012). "Intuition, Subliminal Perception and the Subconscious." Retrieved 12.08.2012, from http://www.evsc.net/home/intuition-subliminal-perception-and-the-subconscious.

Schmauks, D. (2015). Spitze Bemerkungen und schwammige Argumente. Tastsinn und Handhabung in Redewendungen. Tübingen, Stauffenburg.

Schmidthals, K. (2005). Wie das Selbst unser Wissen formt: der Einfluss independenten und interdependenten Selbstwissens auf die Kontextabhängigkeit mentaler Repräsentationen. Berlin, Freie Universität Berlin.

Schuller, B. (2014). Computational paralinguistics. Hoboken, N.J., Wiley.

Schwarz-Friesel, M. (2013). Sprache und Emotion. Tübingen, Francke.

Seufert, G. (1955). Farbnamenlexikon von A-Z. Göttingen, Musterschmidt.

Shute, V. and B. Towle (2003). "Adaptive E-Learning." Educational Psychologist 38(2): 105-114.

Shute, V., M. Ventura, M. Bauer and D. Zapata-Rivera (2008). Melding the Power of Serious Games and Embedded Assessment to Monitor and Foster Learning: Flow and Grow. The Social Science of Serious Games: Theories and Applications. Ritterfeld, U., M. Cody and P. Vorderer. Philadelphia, Routledge/LEA.

Singh, J. and V. Gupta (2017). "A systematic review of text stemming techniques." Artificial Intelligence Review 48(2): 157-217.

Šisler, V. (2016). "Contested Memories of War in Czechoslovakia 38-89: Assassination: Designing a Serious Game on Contemporary History." Game Studies 16(2).

Sommeregger, P. and G. Kellner (2012). Brief Guidelines for Educational Adventure Games Creation (EAGC). Digital Game and Intelligent Toy Enhanced Learning (DIGITEL), 2012 IEEE Fourth International Conference on.

Stieglitz, S., L. Dang-Xuan, A. Bruns and C. Neuberger (2014). "Social Media Analytics." Wirtschaftsinformatik 56(2): 101-109.

Stone, H., R. N. Bleibaum and H. A. Thomas (2012). Sensory evaluation practices. London [u.a.], Elsevier Acad. Press.

Taboada, M. (2016). "Sentiment Analysis: An Overview from Linguistics." Annual Review of Linguistics 2(1): 325-347.

Tan, P.-H., S.-W. Ling and C.-Y. Ting (2007). Adaptive digital game-based learning framework. Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts. Perth, ACM: 142-146.

Tausczik, Y. R. and J. W. Pennebaker (2010). "The psychological meaning of words: LIWC and computerized text analysis methods." Journal of Language and Social Psychology 29(1): 24-54.

Thelwall, M. and K. Buckley (2013). "Topic-based sentiment analysis for the social web: The role of mood and issue-related words." Journal of the American Society for Information Science and Technology 64(8): 1608-1617.

Truong, H. M. (2016). "Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities." Computers in Human Behavior 55(Part B): 1185-1193.

Tsianos, N., P. Germanakos, Z. Lekkas, C. Mourlas and G. Samaras (2010). "Working Memory Span and E-Learning: The Effect of Personalization Techniques on Learners’ Performance." User Modeling, Adaptation, and Personalization: 64-74.

Tsianos, N., Z. Lekkas, P. Germanakos, C. Mourlas and G. Samaras (2008). User-Centric Profiling on the Basis of Cognitive and Emotional Characteristics: An Empirical Study. Adaptive Hypermedia and Adaptive Web-Based Systems. Nejdl, W., J. Kay, P. Pu and E. Herder, Springer Berlin Heidelberg: 214-223.

Tune, S., M. Schlesewsky, A. Nagels, S. L. Small and I. Bornkessel-Schlesewsky (2016). "Sentence understanding depends on contextual use of semantic and real world knowledge." NeuroImage 136: 10-25.

Underwood, J. S., S. Kruse and P. Jakl (2010). Moving to the Next Level: Designing Embedded Assessments into Educational Games. Design and Implementation of Educational Games: Theoretical and Practical Perspectives, IGI Global: 126-140.

Vaishnavi, V. K. and W. Kuechler (2015). Design science research methods and patterns. Innovating information and communication technology. Boca Raton, CRC Press.

Vester, F. (2011). Denken Lernen Vergessen. Was geht in unserem Kopf vor, wie lernt das Gehirn, und wann lässt es uns im Stich? München, dtv.

Vliegen, M. (1988). Verben der auditiven Wahrnehmung im Deutschen. Eine semantisch-syntaktische Analyse. Tübingen, Narr

Volkova, S., Y. Bachrach and B. v. Durme (2016). Mining User Interests to Predict Perceived Psycho-Demographic Traits on Twitter. IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService). Oxford, United Kingdom, IEEE.

Waila, P., V. K. Singh and M. K. Singh (2013). Blog text analysis using topic modeling, named entity recognition and sentiment classifier combine International Conference on Advances in Computing, Communications and Informatics (ICACCI). Mysore, India: 1166-1171.

Wang, C. C. and C. C. Chen (2010). "Electronic commerce research in latest decade: A literature review." International Journal of Electronic Commerce 1(1): 1-14.

Wäsch, S. (2005). Wortschatz der Lebensmittelsensorik im Deutschen. Ein Pilotversuch. Semantik der Sinne. Linke, A. and J. Nuessli. Zürich, Labor für Lebensmittelchemie und -technologie: 37-44.

Weisgerber, L. (1928). "Der Geruchssinn in unseren Sprachen." Indogermanische Forschungen 156: 121-150.

Weisgerber, L. (1929). "Adjektivische und verbale Auffassung der Gesichtsempfindungen." Wörter und Sachen 12: 197-226.

White, R., I. Ruthven and J. Jose (2002). The Use of Implicit Evidence for Relevance Feedback in Web Retrieval. Advances in Information Retrieval. Crestani, F., M. Girolami and C. van Rijsbergen. Berlin Springer: 449-479.

Wiebe, J., T. Wilson, R. Bruce, M. Bell and M. Martin (2004). "Learning Subjective Language." Computational Linguistics 30(3): 277-308.

Wieringa, R. (2014). Design science methodology for information systems and software engineering. Berlin, Springer.

Winkler, S., Ed. (2015). Ambiguity: Language and Communication. Berlin, De Gruyter.

Wolfe, J. M., K. R. Kluender, D. M. Levi, L. M. Bartoshuk, R. S. Herz, R. L. Klatzky and S. J. Lederman (2006). Sensation and perception. Sunderland, Sinauer.

Woo, H. and H. J. Ahn (2015). "Big Five personality and different meanings of happiness of consumers." Economics & Sociology 8(3): 145-154.

Yadollahi, A., A. G. Shahraki and O. R. Zaiane (2017). "Current State of Text Sentiment Analysis from Opinion to Emotion Mining." ACM Computational Surveys 50(2): 1-33.

Yarkoni, T. (2010). "Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers." Journal of research in personality 44(3): 363-373.

Young, K. A. (2005). "Direct from the source: the value of 'think aloud' data in understanding learning." The Journal of Educational Enquiry 6(1): 19-33.

Zeevat, H. (2014). Language production and interpretation. Leiden, Brill.

Zemb, J. M. (1984). Vergleichende Grammatik Französisch–Deutsch. Part 1: Comparaison de deux systèmes. Part 2: L'économie de la langue et le jeu de la parole. Mannheim, Duden.

Zhai, C. and S. Massung (2016). Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining. New York, Morgan & Claypool.

Zimmer, K. (2006). Die Sprache der Parfüms. Wortschöpfungen und Wortfeldübertragungen in der fach- und werbesprachlichen Darstellung von Düften. Stuttgart. Master Thesis.

Zwick, W. R. and W. F. Velicer (1986). "Comparison of five rules for determining the number of components to retain." Psychological Bulletin 99: 432-442.

Appendix

List of Figures

Figure 1 The Information Systems Research Framework [Hevner et al. 2004] 15
Figure 2 Our Research Methodology 16
Figure 3 Contribution Classification according to the Knowledge Contribution Framework 17
Figure 4 Perceptual Limitations During Information Processing [Schindling 2012] 25
Figure 5 The Perception-Action Loop [Ernst and Bülthoff 2004] 26
Figure 6 Factor Analysis Scree Plot: Eigenwert (x-scale) per Number of Factors (y-Scale) 41
Figure 7 PPQ Test Items: Frequency (x-scale) per Likert-Scale option 1-5 (y-scale) 42
Figure 8 Co-occurrences of interests 45
Figure 9 Correlations between perceptual preferences and interests 46
Figure 10 Some German words that include the letters seh [see] 56
Figure 11 Subsets of search results 57
Figure 12 Parameterization of the algorithms 60
Figure 13 Decision problem 61
Figure 14 Word fragmentation 65
Figure 15 Class diagram: Database interface 66
Figure 16 The STEM algorithm and its components 73
Figure 17 Word Components: Prefixes 75
Figure 18 Word Components: Preendings 76
Figure 19 Word Components: Endings 76
Figure 20 Word Components: Word joints 76
Figure 21 Flowchart diagram for the Surrounding letters check 78
Figure 22 Execution Example: Wangenrötung 78
Figure 23 Execution Example: Schildkröten 79
Figure 24 Hypotheses A–C and their testing 83
Figure 25 Full-text and VAKOG similarity in our samples 89
Figure 26 Distribution of full-text and VAKOG similarity within the samples 90
Figure 27 The EAGC Process: Stages and Dependencies 103
Figure 28 Screenshots of the educational adventure game “Save the city” 109
Figure 29 Screenshots of the four GETOLS minigames. 110
Figure 30 The GETOLS Testing Procedure 111

List of Tables

Table 1 The 46 PPQ Test Item Candidates ____ 37
Table 2 PPQ Test Items ____ 38
Table 3 PPQ Test Items: Not Validated English Translation ____ 38
Table 4 List of 27 Interests ____ 39
Table 5 PPQ Test Items ____ 43
Table 6 Some words Based on the stem seh [see] and its realizations ____ 55
Table 7 Part of speech (PoS) realization types ____ 61
Table 8 Word fragment classification ____ 65
Table 9 Lexemes chosen for evaluation ____ 68
Table 10 Algorithm evaluation results ____ 69
Table 11 Reduction criteria per subset ____ 86
Table 12 E30 Forum Corpus: Reduction criteria and the resulting amount of data per subset_ 88
Table 13 The distribution of sensory expression in the E30 corpus ____ 89
Table 14 Suitability of Different Testing Methods For GETOLS ____ 105
Table 15 Validation of GETOLS Video Use for Kinesthetic Testing ____ 106
Table 16 Comparison of Methods: Results by Order ____ 113
Table 17 Comparison of Methods: Results by Scores ____ 114
Table 18 Comparison of Methods: Score Differences ____ 115

Acknowledgements

This PhD thesis would be different, and for sure poorer, without the inspiration and support I got from the people around me.

First of all, I want to thank those who contributed to my research, who took the time to dive deep into my concepts, to think of possible extensions and sometimes to even realize parts together with me. My special thanks goes to my supervisor Prof. Dieter Merkl, who accompanied me during my whole PhD time and gave me valuable advice, and to Prof. Bettina Berendt, who hosted me as a guest researcher at the KU Leuven and motivated me to go further with my research. I would like to thank my colleagues and the professors at the Electronic Commerce Research Group at the Institute of Software Technology and Interactive Systems, Vienna University of Technology, and especially the head of the group, Prof. Hannes Werthner, for constructive exchange and fruitful discussions. Thanks to Marcel Berthold for the enriching cooperation and to my students Johannes Grünauer, Paul Sommeregger, and Andreas Weißenbacher for going beyond what was necessary to only obtain your grades. Thanks a lot to Julia Richling for giving us full access to your forum corpus.

This work would not have been possible without financial support of the Vienna University of Technology and the Austrian Federal Ministry of Science and Research, jointly funding the fFORTE WIT Program. A great thanks to Sabine Cirtek and Vera Kuzmits for all the support of interdisciplinary and interfaculty exchange.

Thanks to the international research community for awarding my work with best paper and best poster awards. In a phase where I was desperately looking for one single scientific context that I could call my own, this highly appreciated feedback made me believe again in the potential of my interdisciplinary approach.

Thanks to all my friends for taking interest in my PhD, and for sharing my despair and my joys during setbacks and achievements. A huge thank you goes to Maria Salamon, for being a great discussion partner, for your understanding and your support during the whole time. I'm very grateful to my family and especially to my parents, Barbara Fuchshuber und Georg Kellner, for supporting my studies and for believing in me.

I owe my gratitude to all of you.