Predicting Expertise from eye movements in reading

Project report

Lena Jäger

June 2, 2016

Abstract

4 groups of participants (advanced/beginner level students of physics and biology) read a series of short texts taken from textbooks of physics and biology while their eye movements were monitored (each participant reads all texts). Their text comprehension as well as their background knowledge in the topics presented in the texts were assessed by multiple-choice comprehension questions. The readers' eye movements in combination with various lexical und supra-lexical features of the texts were modelled to predict their reading comprehension and their background knowledge.

1 Introduction: Aim of this research

It is widely known that reading comprehension influences a reader's eye movements on the lexical, syntactic, semantic and discourse level of the language [15, 12, 20, 19]. Moreover, it has been shown that eye movements allow to identify a reader with a relatively high accuracy [13, 11]. The aim of this research is to develop a model that is able to a predict a reader's higher level text comprehension as well as his background knowledge in the topic presented in a text that is being read from his eye movements and their interaction with the properties of the text.

2 Data collection

Advanced and first-semester students of physics or biology each read a series of texts on biology and on physics. This resulted in a 2×2 fully-crossed factorial design with the reader's expertise (advanced vs beginner) and the reader's major (physics vs biology) as factors. The text comprehension as well as the reader's background knowledge were assessed by presenting multiple-choice questions after each text.

2.1 Stimuli

2.1.1 Text selection and editing

A total of 12 texts were selected from various German language physics and biology textbooks [5, 6, 7, 1, 18, 9, 2]. The texts deal with various topics usually covered in standard introductory university level lectures on physics (6 texts) and biology (6 texts). The texts were chosen such that university level knowledge is necessary to fully understand their contents. The fulfilment of this criterion was assessed by three experts/non-experts who read and judged the texts' difficulty level. The original texts were edited in a way that each text was approximately 150 words (min: 126 words; max: 180 words; mean: 158 words)[1] long. The stimuli were edited such that they contained only plain text; mathematical formulas, figures and tables as well as references to them were removed and, if necessary, the contents of the texts were adjusted to account for their removal.

[1] min: 1030 characters; max: 1281 characters; mean: 1196 characters

2.1.2 Text features

2.1.2.1 Part-of-speech-tags

Each text was manually part-of-speech tagged according to the Stuttgart-Tübingen-Tagset (STTS) [16].

2.1.2.2 Corpus-based lexical-level features

For each word, several corpus-based features commonly used in reading research were semi-automatically (see `features.py`) extracted from the lexical database dlexDB [8, 10], which is based on the reference corpus underlying the Digital Dictionary of the German Language (DWDS) corpus [4]. Please refer to Table 1 for the definitions of the various features.

Feature	Definition
Linguistic representations	
type	orthographical representation of a word as found in the corpus
lemma	headword, i.e., an uninflected form that may or may not occur in the corpus itself
syllables	syllables of which the word consists
Word length measures	
type length (number of characters)	number of characters of a type
type length (number of syllables)	number of syllables of a type
lemma length (number of characters)	number of characters of a lemma
Frequency measures	
annotated type frequency	number of occurrences of a unique combination of a type, its STTS tag and its lemma in the corpus (per mio tokens)
type frequency	number of occurrences of a type in the corpus (per mio tokens)
lemma frequency	total number of occurrences of types associated with this lemma in the corpus (per mio tokens)
document frequency	the number of documents with at least one occurrence of this type (per 10.000 documents)
sentence frequency	number of sentences with at least one occurrence of this type (per 100.000 sentences)
cumulative syllable corpus frequency	number of occurrences in the syllabified representation of the corpus (per mio tokens)
cumulative syllable lexicon frequency	number of occurrences in the syllabified list of types (per mio types)
cumulative character corpus frequency	cumulative corpus frequency of all characters contained in this type (per mio tokens)
cumulative character lexicon frequency	cumulative lexicon frequency of all characters contained in this type (per mio types)
cumulative character bigram corpus frequency	cumulative corpus frequency of all character bigrams contained in this type (per mio tokens)
cumulative character bigram lexicon frequency	cumulative lexicon frequency of all character bigrams contained in this type (per mio types)
cumulative character trigram corpus frequency	cumulative corpus frequency of all character trigrams contained in this type (per mio tokens)
cumulative character trigram lexicon frequency	cumulative lexicon frequency of all character trigrams contained in this type (per mio types)
initial letter frequency	cumulative frequency of all types sharing the same initial letter (per mio tokens)
initial bigram frequency	cumulative frequency of all types sharing the same initial character bigram (per mio tokens)
initial trigram frequency	cumulative frequency of all types sharing the same initial character trigram (per mio tokens)
average conditional probability (in bigrams)	conditional probability of the bigram, given the occurrence of its first component, averaged across all character positions (computed on the basis of the annotated type information)
average conditional probability (in trigrams)	conditional probability of the trigram, given the occurrence of its first component, averaged across all character positions (computed on the basis of the annotated type information)
familiarity	cumulative frequency of all types of the same length sharing the same initial trigram
regularity	the number of types of the same length sharing the same initial trigram
Neighborhood measures	
cumulative frequency of higher frequency neighbors (Coltheart)	cumulative frequency of all higher frequency orthographic neighbors according to the definition of [3]
count of of higher frequency neighbors	number of higher frequency orthographic neighbors according to the definition of [3]
cumulative frequency of all neighbors (Coltheart)	cumulative frequency of all orthographic neighbors according to the definition of [3]
count of all neighbors (Coltheart)	number of orthographic neighbors according to the definition of [3]
cumulative frequency of higher frequency neighbors (Levenshtein)	cumulative frequency of all higher frequency orthographic neighbors according to the definition of [14]
count of of higher frequency neighbors (Levenshtein)	number of higher frequency orthographic neighbors according to the definition of [14]
cumulative frequency of all neighbors (Levenshtein)	cumulative frequency of all orthographic neighbors according to the definition of [14]
count of all neighbors (Levenshtein)	number of orthographic neighbors according to the definition of [14]

Table 1: Definition of the corpus-based features extracted from dlexDB.

Feature	Definition
Lexical and orthographic features	
technical term	codes whether the word is a technical term and for technical terms whether it is commonly used in non-technical contexts or only in technical texts (ternary variable)
is abbreviation	codes whether the token is an abbreviation
contains abbreviation	codes whether the token contains an abbreviation
symbol	word contains a symbol (e.g., z-Richtung, +Ende; β-D-Glucose)
hyph	word contains at least one hyphen that is *not* STTS-tagged as TRUNC (e.g., DNA-Fragment, z-Richtung, β-D-Glucose)
Linear position information	
word index in text	position of the word within the current text, irrespective of sentences coded as integer
word index in sentence	position of the word within the current sentence coded as integer
sentence index	position of the sentence to which the word belongs within the current text coded as integer
Punctuation	
STTS punctuation before	STTS-tag of the punctuation that precedes the word
STTS punctuation after	STTS-tag of the punctuation that follows the word
quote	word is (part of an expression that is) in quotes
parentheses	word is (part of an expression that is) in parentheses
Syntactic features	
clause begin	first word of a new clause
sentence begin	first word of a sentence

Table 2: Definition of other word- and sentence-level features.

The entries extracted from the dlexDB database were semi-automatically corrected. First, the type-to-lemma mapping was ambiguous in many cases. The correct lemma was selected manually. Second, there was a non-negligible number of entries with coding mistakes in the database (e.g., incorrect lemma etc.). If possible, this was corrected manually, otherwise the obviously wrong numbers were re-coded as missing values.

2.1.2.3 Other word and sentence-level features

In addition to the PoSTags and the corpus-based features, each word was coded for a series of other features that might influence the reading behavior on this word. These features are defined in Table 2.

2.1.3 Comprehension questions

For each text, three text comprehension questions and three background questions were created. The text comprehension questions required a thorough understanding of the text, but did not require any additional background knowledge. The background questions, in contrast, tested the general knowledge in the topic presented in the text and hence required background knowledge.

2.2 Participants

2.2.1 Selection criteria

All participants were native speakers of German with normal or corrected-to-normal[2] vision. They were either students of biology or of physics in either their first semester of the BSc program or advanced students currently

[2]Participants with soft lenses were excluded from participation as soft lenses often pose serious challenges for the calibration of the eye-tracker.

attending a MSc or PhD program. Participants were requested to not have consumed any alcohol the day of the experiment.

2.2.2 Participants' personal data

The field of studies (including area specialization if applicable), the current semester of studies, gender, age, handedness, whether the participant was wearing contact lenses or glasses, hours of sleep the night before the experiment, alcohol consumption within 24h hours before the experiment, whether or not the participant had grown up bilingually, and the state (Bundesland) where the German language was acquired were recorded.

2.3 Eye movements and text comprehension

2.3.1 Technical set-up

Participant's eye movements (right eye monocular tracking) were recorded with an Eyelink 1000 tracker manufactured by SR Research at a sampling rate of 1000 Hz using a desktop mounted camera system with a 35 mm lens. The experimental presentation and the communication between presentation computer and eye-tracker was implemented using Experiment Builder software provided by SR Research.

The participant was seated at a hight-adjustable table to ensure a constant eye-to-screen distance across participants. The participant's head was stabilized using a forehead and chin rest. The texts were presented on a 22 inch monitor with a resolution of 1680×1050 pixels. The eye-to-screen distance measured 61 cm and the eye-to-camera distance was 65 cm.

The texts were presented in a monospaced white font (Courier, font size 18) on a black background. The reason for choosing a black background was the rather long duration of the experiment. A bright background color would strain the participants' eyes and potentially lead to wet eyes which has a negative impact on calibration accuracy.[3] As a response pad a Cedrus button box was used. For the comprehension questions the buttons to be pressed with the left and right index fingers and middle fingers were used; as 'continue'-button, the two thumb buttons were used.

2.3.2 Procedure

Before the start of the experiment, the participants were informed about the experimental procedure (for the precise wording of the participant briefing see Figure 3 in the appendix) and signed an ethical consent form. After the set-up and calibration of the camera, the participant first read one practice text followed by six practice questions to get familiar with the experimental procedure. The twelve experimental texts were presented in randomized order (separate randomization for each participant), see `umwandlung.py` for the formatting and randomization of the stimuli. Each experimental trial began with the header of the following text being presented. The participant had to press a button to continue to the text. All texts fit onto one screen. There were no restrictions regarding the time spent on reading one text. After having finished reading the text, the participant had to press a button go continue to the comprehension questions. Each comprehension question was presented on a separate screen together with 4 multiple choice options of which the participant had to select one by pressing the respective button on the button box. It was not possible to go back to the text or previous questions nor was it possible to undo an answer. The buttons on the button box were randomly assigned to an answer option for each participant. The three text comprehension questions always preceded the background questions; the three questions within each type were randomized for each participant. Participants were informed that some of the questions required background knowledge, however they were not informed which ones.

If necessary, recalibrations were performed before the beginning of a new trial. Participants were allowed to take a break before the beginning of a new trial. To ensure a good calibration quality, female participants were asked to remove mascara before the start of the experiment. The total duration of the experiment was approximately 90 minutes.

[3]The dark background color led to an increased diameter of the participants' pupils, which also rendered the calibration procedure difficult. The solution was to arrange the lab such that natural light illuminated the room in addition to the artificial light.

2.3.3 Participant payment

Each participant received 10 EUR plus a performance-based bonus of up to 10 EUR. Beginners received the full 20 EUR when they answered 31 or more of the 72 questions correctly, whereas experts received the full 20EUR when they answered at least 45 questions correctly. For the precise payment scale see Table 4 in the appendix.

3 Modeling

3.1 Eye movements pre-processing

3.1.1 Computation of fixations

Fixations were computed from the raw data using the Eyelink Data Viewer software package provided by SR Research with the default parameter settings [17].

3.1.2 Computation of word-level reading measures

From the fixation data, various reading measures commonly used in reading research were computed (see `fix2roi.py` and `MergeData.ipynb`). Each word (defined by the surrounding white spaces) was considered one region of interest for all measures except for landing position which was based on the characters within a word. Fixations on the punctuation marks were considered to belong to the preceding word by default and to the following word in case of opening parentheses or opening quotation marks. Definitions of the various measures are provided in Table 3. Note that several of these measures are linearly dependent.

Measure	Abbreviation	Definition
Continuous measures (in ms)		
first-fixation duration	FFD	duration of the first fixation on a word if this word is fixated in first-pass reading, otherwise 0
first duration	FD	duration of the first fixation on a word (identical to FFD if not skipped in the first-pass)
first-pass reading time	FPRT	sum of the durations of all first-pass fixations on a word (0 if the word was skipped in the first-pass)
single-fixation duration	SFD	duration of the only first-pass fixation on a word, 0 if the word was skipped or more than one fixations occurred in the first-pass (equals FFD in case of a single first-pass fixation)
first-reading time	FRT	sum of the duration of all fixations from first fixating the word (independent if the first fixations occurs in first-pass reading) until leaving the word for the first time (equals FPRT in case the word was fixated in the first-pass)
total-fixation time	TFT	sum of all fixations on a word (FPRT+RRT)
re-reading time	RRT	sum of the durations of all fixations on a word that do not belong to the first-pass (TFT−FPRT)
inclusive regression-path duration	RPD_inc	sum of all fixation durations starting from the first first-pass fixation on a word until fixation a word to the right of this word (including all regressive fixations on previous words), 0 if the word was not fixated in the first-pass (RPD_exc+RBRT)
exclusive regression-path duration	RPD_exc	sum of all fixation durations after initiating a first-pass regression from a word until fixating a word to the right of this word, without counting fixations on the word itself (RPD_inc−RBRT)
right-bounded reading time	RBRT	sum of all fixation durations on a word until a word to the right of this word is fixated (RPD_inc−RPD_exc).
Binary measures		
fixation	Fix	1 if the word was fixated, otherwise 0 (FPF `or` RR)
first-pass fixation	FPF	1 if the word was fixated in the first-pass, otherwise 0
first-pass regression	FPReg	1 if a regression was initiated in the first-pass reading of the word, otherwise 0 (`sign`(RPD_exc))
re-reading	RR	1 if the word was fixated after the first-pass reading, otherwise 0 (`sign`(RRT))
Ordinal measures		
landing position	LP	position of the first saccade on the word expressed by ordinal position of the fixated character
incoming saccade length	SL_in	length of the saccade that leads to first fixation on a word in number of words; positive sign if the saccade is a progressive one, negative sign if it is a regression
outgoing saccade length	SL_out	length of the first saccade that leaves the word in number of words; positive sign if the saccade is a progressive one, negative sign if it is a regression; 0 if the word is never fixated
total count of incoming regressions	TRC_out	total number of regressive saccades initiated from this word
total count of outgoing regressions	TRC_in	total number of regressive saccades landing on this word

Table 3: Definition of the word-level reading measures.

Appendix

Participant briefing

Herzlich Willkommen zu unserer Eyetracking-Lesestudie!
Wir danken dir ganz herzlich für deine Teilnahme!

In diesem Experiment wirst du insgesamt 12 kurze Texte zu physikalischen oder biologischen Themen lesen, die maximal eine Bildschirmlänge lang sind. Nach jedem Text beantwortest du inhaltliche Multiple-Choice-Fragen per Tastendruck. Während du liest, nimmt der Eyetracker die Bewegungen deiner Augen auf. Du bist berechtigt, das Experiment jederzeit ohne Angabe von Gründen abzubrechen. Bitte lies zunächst aufmerksam die folgenden Erläuterungen durch.

Der Ablauf für jeden Text ist wie folgt:

1) An der Stelle, wo das erste Wort des Textes erscheinen wird, wird zunächst ein grauer Kalibrierungspunkt mit einem **kleinen schwarzen Punkt in der Mitte** erscheinen. Sobald du auf diesen winzigen **schwarzen Mittelpunkt** schaust, wird automatisch der Text angezeigt.
2) Lies den Text ganz natürlich (ohne Stimme). **Wichtig ist, dass du versuchst, den Text inhaltlich zu verstehen.**
3) Sobald du den Text fertiggelesen hast, schaue bitte auf den **runden grünen Aufkleber** in der unteren rechten Ecke des Bildschirms und drücke eine der beiden "weiter"-Tasten.
4) Nun erscheinen nacheinander sechs Fragen zu dem Text. Alle Fragen sind Multiple-Choice-Fragen mit jeweils vier Antwortmöglichkeiten. **Es ist immer nur eine Antwort richtig.** Zum Auswählen deiner Antwort drückst du bitte die Taste mit der entsprechenden Nummer auf der Button-Box. Die Fragen sind teilweise recht schwierig und manche der Fragen prüfen auch Fachwissen, das über den eigentlichen Inhalt des Textes hinausgeht. Es ist sehr wichtig, dass du versuchst, alle Fragen trotzdem so gut wie möglich zu beantworten. Wenn du dir nicht sicher bist, wähle die Antwort, die du für am plausibelsten hälst.

Bevor das eigentliche Experiment beginnt, wirst du einen Übungstext mit Fragen sehen, damit du dich an den Ablauf gewöhnen kannst. Thema des Übungstexts ist der Experimentablauf. Sollte noch etwas unklar sein, kannst du nach dem Übungstext die Versuchsleiterin fragen. Während des eigentlichen Experiments sollten nach Möglichkeit keine Fragen mehr gestellt werden.
Bitte rücke den Stuhl nah an den Tisch heran und nimm eine **aufrechte Haltung** ein. Lege dein Kinn auf die dafür vorgesehene Halterung und lehne die Stirn an den Rahmen. Die Tischhöhe kann an deine Größe angepasst werden. Die Versuchsleiterin wird dir helfen, Tischhöhe und Kinnstütze anzupassen. Bitte lege deine Finger auf die Tasten der Button-Box, so dass die beiden Daumen auf den "weiter-Tasten" liegen und der linke Mittelfinger auf der 1, der linke Zeigefinger auf der 2, der rechte Zeigefinger auf der 3 und der rechte Mittelfinger auf der 4 liegt.

Wenn du deine Sitzposition gefunden hast, wird der Eyetracker auf deine Augen kalibriert. Auf dem Bildschirm werden graue Punkte erscheinen, in deren Mitte ein **winziger schwarzer Punkt** ist. Folge mit deinem Blick diesen winzigen schwarzen Punkten. Bitte fixiere jeden Punkt solange, bis er verschwindet. Bitte **vermeide Kopfbewegungen** während und nach der Kalibrierung. Solltest du den Kopf bewegen, muss die Kalibrierung leider wiederholt werden.

Du kannst immer dann eine **Pause** machen, nachdem du einen Text fertig gelesen und die Fragen dazu beantwortet hast. Sag einfach der Versuchsleiterin Bescheid! Falls du dich aus irgendeinem Grund schlecht fühlen solltest, kannst du natürlich jederzeit unterbrechen bzw. abbrechen.

Vergütung: Du bekommst nach Beenden des Experiments mindestens 10EUR ausbezahlt. Darüber hinaus kannst du dir einen Bonus von bis zu 10EUR zusätzlich verdienen: Je mehr Fragen du richtig beantwortest, desto höher ist dein Bonus. Liegt deine Antwortgenauigkeit auf dem Zufallsniveau, bekommst du keinen Bonus. Den vollen Bonus bekommen Studienanfänger, wenn sie mindestens 43% aller Fragen richtig beantworten. Fortgeschrittene bekommen den vollen Bonus, wenn sie mindestens 63% aller Fragen richtig beantworten. Die Bezahlung erfolgt direkt nach dem Experiment in bar.

Viel Spaß!

Figure 3: Participant briefing read by each participant before the experiment.

Participant payment

Correct responses	Payment beginner (€)	Payment expert (€)
0–18	10	10
19–21	12	10
22–24	14	12
25–27	16	12
28–30	18	14
31–33	20	14
34–36	20	16
37–39	20	16
40–42	20	18
43–45	20	18
45–72	20	20

Table 4: Participant payment including the performance-based bonus paid to beginners (1st semester students) and experts (MSc/PhD students) by number of correctly answered questions.