

Authentic Text ICALL (ATICALL) Exercise Generation & Information Retrieval for Language Learning

Niels Ott and Detmar Meurers

Berlin, October 16, 2009

ATICALL
Niels Ott and Detmar Meurers

Introduction

Pedagogical grounding

The WERTI system

Modeling FLT practice

Progression in WERTI

Examples and Target types

Realizing proposal

Creating ex. progression

Prototype realization

Some challenges

IR4LL (Ott 2009)

Measuring Text Difficulty

Readability Formulas

Lexical Frequency Profiles

Syntactic Complexity

Textbook structures

A Search Engine Prototype

Information Retrieval

Adding Text Difficulty

Search Evaluation

Summary

Appendix

Prototype screenshots

IR4CALL NLP pipeline

Introduction

- ▶ The use of NLP in ICALL has primarily centered on diagnosing learner errors and, more recently, testing and assessment.
- ▶ Idea: Explore how NLP technology can support other aspects of second language learning.
- ▶ Our specific focus: What can NLP contribute to **awareness of language forms and rules**, an important component of adult second language acquisition?
 - ▶ WERTI: Automatic generation of language awareness activities based on real-world texts.
 - ▶ IR4LL: Retrieval of authentic texts at the appropriate level for language learners

ATICALL
Niels Ott and Detmar Meurers

Introduction

Pedagogical grounding

The WERTI system

Modeling FLT practice

Progression in WERTI

Examples and Target types

Realizing proposal

Creating ex. progression

Prototype realization

Some challenges

IR4LL (Ott 2009)

Measuring Text Difficulty

Readability Formulas

Lexical Frequency Profiles

Syntactic Complexity

Textbook structures

A Search Engine Prototype

Information Retrieval

Adding Text Difficulty

Search Evaluation

Summary

Appendix

Prototype screenshots

IR4CALL NLP pipeline

Pedagogical grounding of our research Awareness

Awareness (Schmidt 1995):

▶ Noticing

- ▶ "conscious registration of an event"
- ▶ low level of awareness
- ▶ implicit learning

E.g.: noticing that sometimes speakers of Spanish omit the subject pronoun

▶ Understanding

- ▶ "recognition of a general principle, rule or pattern"
- ▶ higher level of awareness
- ▶ explicit learning
- ▶ generalization can be internally generated or externally provided

E.g. understanding that Spanish is a pro-drop language

ATICALL
Niels Ott and Detmar Meurers

Introduction

Pedagogical grounding

The WERTI system

Modeling FLT practice

Progression in WERTI

Examples and Target types

Realizing proposal

Creating ex. progression

Prototype realization

Some challenges

IR4LL (Ott 2009)

Measuring Text Difficulty

Readability Formulas

Lexical Frequency Profiles

Syntactic Complexity

Textbook structures

A Search Engine Prototype

Information Retrieval

Adding Text Difficulty

Search Evaluation

Summary

Appendix

Prototype screenshots

IR4CALL NLP pipeline

Pedagogical grounding of our research The role of awareness

▶ Research on awareness shows:

- ▶ There is no learning without noticing.
- ▶ Awareness without input is not sufficient.
- ▶ "Learning takes place within the learner's mind and cannot be completely engineered by teachers or syllabus designers."
- ▶ One can only provide opportunities for developing learner awareness.

⇒ Consequences:

- ▶ Learners have to be exposed to linguistic features to acquire them.
- ▶ Learners have to notice those features.
- ▶ Tools presenting such linguistic features in a contextualized way, allowing for student interaction, can be helpful.

ATICALL
Niels Ott and Detmar Meurers

Introduction

Pedagogical grounding

The WERTI system

Modeling FLT practice

Progression in WERTI

Examples and Target types

Realizing proposal

Creating ex. progression

Prototype realization

Some challenges

IR4LL (Ott 2009)

Measuring Text Difficulty

Readability Formulas

Lexical Frequency Profiles

Syntactic Complexity

Textbook structures

A Search Engine Prototype

Information Retrieval

Adding Text Difficulty

Search Evaluation

Summary

Appendix

Prototype screenshots

IR4CALL NLP pipeline

Pedagogical grounding of our research

Linguistic information and how it is conveyed

- ▶ A wide range of linguistic features can be relevant for awareness, incl. morphological, syntactic, semantic, and pragmatic information (cf. Schmidt 1995, p. 30).
 - ▶ Linguistic information can be conveyed to the learner
 - ▶ using **explicit** linguistic terminology/representations, e.g.:
 - ▶ parts of speech
 - ▶ verbal tense, mood and aspect
 - ▶ sentence classification
 - ▶ syntactic analyses (shown as trees or sentence diagrams)
 - ▶ using **implicit** presentation, e.g.:
 - ▶ coloring, underlining, moving, etc
 - ▶ pointing to correct or incorrect uses
- ⇒ Awareness activities can include both implicit and explicit presentation of linguistic features.

ATICALL
Nelle Oti and Damar Meurers

Introduction
Pedagogical grounding

The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulae
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype screenshots
IR/ECALL NLP pipeline

Modeling FLT practice

- ▶ A common pedagogical practice in FLT moves from target language presentation, to practice, on to production.
- ▶ Proposal: Create sequences of linguistic awareness activities following the initial stages of such a progression:
 - I. Receptive presentation
 - II. Productive presentation
 - III. Controlled practice
- ▶ What makes this idea interesting?
 - ▶ NLP technology can identify certain relevant linguistic categories and forms in real-life texts.
 - ▶ The contents of these texts can be selected by the learners based on their interests.
 - ▶ The sentences turned into exercises can remain fully contextualized as part of the text selected by learner.
 - ▶ Automatic feedback for the activities is feasible since the original text is known.

ATICALL
Nelle Oti and Damar Meurers

Introduction
Pedagogical grounding

The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulae
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype screenshots
IR/ECALL NLP pipeline

The activity progression in WERTI

Using real world web-based texts (such as news articles) we provide a progression of activities:

Step 1. Receptive presentation

Ex. The system **colors** examples of targeted items.

Step 2. Productive presentation

Ex. The learner is asked to **find and mouse-click** all tokens of the targeted category. The system shows correct picks in green, incorrect ones in red.

Step 3. Controlled practice

Ex. The learner is asked to

- ▶ **reorder** words/phrases given (scrambled) list
- ▶ complete **fill-in-the-blank** (FIB) slots
- ▶ created for tokens of targeted category
- ▶ given some information, where needed (e.g., stems)

ATICALL
Nelle Oti and Damar Meurers

Introduction
Pedagogical grounding

The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulae
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype screenshots
IR/ECALL NLP pipeline

Examples and Target types

- ▶ Examples:
 - ▶ FIB Determiners
 - ▶ Colored Gerunds
- ▶ Types of targets:
 - ▶ Lexical targets:
 - ▶ prepositions
 - ▶ determiners
 - ▶ Lexical form targets with contextual triggers:
 - ▶ gerunds vs. *to*-infinitives
 - ▶ *if* conditionals
 - ▶ tense and aspect
 - ▶ Syntactic targets with discourse context triggers:
 - ▶ active vs. passive

ATICALL
Nelle Oti and Damar Meurers

Introduction
Pedagogical grounding

The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulae
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype screenshots
IR/ECALL NLP pipeline

What is involved in realizing such an approach?

- ▶ Two components can be distinguished:
 1. Obtaining and selecting appropriate texts:
 - ▶ Texts obtained through web search using terms provided by the language learner
 - restrict web to news sites (e.g., Reuters)
 - alternative: specific corpora
 - ▶ Texts could be filtered according to aspects relevant to language learning (text readability, frequency of relevant constructions, etc. → IR4LL discussion below)
 2. Identifying the targets in the selected texts and creating
 - ▶ receptive and productive presentations, and
 - ▶ controlled practice exercises using the texts.
- ▶ We illustrate the approach, focusing on the second component, by showcasing an activity progression targeting prepositions.

ATICALL
Niels Ott and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IR4LL (Ott 2009)
Measuring Text Difficulty
Readability Formulae
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Search Evaluation

Summary

Appendix
Prototype screenshots
IR4CALL NLP pipeline

ERHARDT-KARL
UNIVERSITÄT
TÜBINGEN

9 / 45

Realizing the proposal

Creating an activity sequence

- ▶ The system first annotates the web page text using efficient and robust NLP tools performing
 - ▶ tokenization → tokens
 - ▶ lemmatization → word roots
 - ▶ part-of-speech tagging → lexical categories
 - ▶ morphological analysis → morphological properties
 - ▶ shallow parsing → phrasal categories
- ▶ The language items targeted by the activity are identified using regular expression matching of target and contextual items in the annotated text.
- ▶ The nature of the activity determines the complexity of the annotation and the regular expressions required:
 - ▶ Preposition activity: single instances of a lexical category
 - ▶ Tense and aspect: sequences of auxiliaries, inflected forms, and specific lexical items (contextual cues)

ATICALL
Niels Ott and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IR4LL (Ott 2009)
Measuring Text Difficulty
Readability Formulae
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Search Evaluation

Summary

Appendix
Prototype screenshots
IR4CALL NLP pipeline

ERHARDT-KARL
UNIVERSITÄT
TÜBINGEN

10 / 45

Prototype realization

- ▶ Original prototype in Python, integrated into the Apache2 webserver using mod_python, including:
 - ▶ searching in the Reuters site providing news webpages
 - ▶ linguistic annotation using NLTK (Bird & Loper 2004), TreeTagger (Schmid 1994)
- ▶ Recently reimplemented as UIMA-based Java servlet on Tomcat server (Aleks Dimitrov, Ramon Ziai, Niels Ott).
- ▶ The annotated text is mapped into Color, Click, and FIB presentation code (HTML and JavaScript), and fully integrated in the original web page.
- ▶ Only a standard web browser is needed to use the system.
- ▶ We are working on integrating further target patterns and activities. Prototypes available at:
 - ▶ WERTI: <http://purl.org/net/WERTI>
 - ▶ WERTI2: <http://delos.sfs.uni-tuebingen.de:8080/WERTI>

ATICALL
Niels Ott and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IR4LL (Ott 2009)
Measuring Text Difficulty
Readability Formulae
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Search Evaluation

Summary

Appendix
Prototype screenshots
IR4CALL NLP pipeline

ERHARDT-KARL
UNIVERSITÄT
TÜBINGEN

11 / 45

Realizing the proposal

Some challenges

- ▶ Annotation errors:
 - ▶ Statistical NLP tools are efficient and robust
 - ▶ Such tools make errors, e.g., 3–5% for POS tagging.
 - ▶ What impact do such errors have for the envisaged use?
 - ▶ It is known where errors are likely to arise (cf., e.g., Dickinson & Meurers 2003; Dickinson 2005), so one can avoid basing activities on likely error locations.
- ▶ The complexity of real life:
 - ▶ Real-life texts from the web often have
 - ▶ complex structure
 - ▶ mark-up and integrated multimedia
 - ▶ It is nontrivial to combine that web page structure with the activity based on the annotated text base.

ATICALL
Niels Ott and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IR4LL (Ott 2009)
Measuring Text Difficulty
Readability Formulae
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Search Evaluation

Summary

Appendix
Prototype screenshots
IR4CALL NLP pipeline

ERHARDT-KARL
UNIVERSITÄT
TÜBINGEN

12 / 45

Finding texts appropriate for language learners

- ▶ How can one find authentic texts as reading material or for activity generation (e.g., WERTI)?
 - ▶ Such texts should
 - ▶ be in the language of interest
 - ▶ have the appropriate level of complexity for the learner
 - ▶ contain enough good instances of the language patterns and rules targeted by the activities.
 - ▶ How about simply using the web and a standard web search engine (e.g., google)?
 - ▶ Pro: The Web is huge, and up-to-date information on virtually any topic is available.
 - ▶ Cons: Standard search engines are not aware of reading complexity and language patterns.
- ⇒ Create a dedicated search engine for language learning: IR4LL (Ott 2009)

ATICALL
Nelle Ott and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IR4LL (Ott 2009)
Measuring Text Difficulty
Readability Formulas
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary
Appendix
Prototype screenshots
IR4CALL NLP pipeline

ERHARDT-KARL
UNIVERSITÄT
TÜBINGEN

13 / 45

IR4LL Proposal

- ▶ Create a search engine that is aware of variations in text difficulty.
- ▶ Challenges and research questions:
 - ▶ How to measure text difficulty?
 - ▶ Is there enough variety in text difficulty on the web?
 - ▶ Are there enough 'easy' web pages?



ATICALL
Nelle Ott and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IR4LL (Ott 2009)
Measuring Text Difficulty
Readability Formulas
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary
Appendix
Prototype screenshots
IR4CALL NLP pipeline

ERHARDT-KARL
UNIVERSITÄT
TÜBINGEN

14 / 45

Readability and how to measure it

- ▶ **Readability** or **text difficulty**: refers to the understandability or comprehensibility of a text (Klare 1963).
- ▶ The more **reading proficient** the reader, the less readable texts need to be in order to be understood by this reader.
- ▶ Traditional **readability formulas** try to measure the readability on a scale, e.g. the U.S. grade level scale.

ATICALL
Nelle Ott and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IR4LL (Ott 2009)
Measuring Text Difficulty
Readability Formulas
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary
Appendix
Prototype screenshots
IR4CALL NLP pipeline

ERHARDT-KARL
UNIVERSITÄT
TÜBINGEN

15 / 45

U.S. grade level scale

Scale based on Gunning (1968, p. 40):

Grade Level	Named Grade
17	College graduate
16	senior
15	junior
14	sophomore
13	freshman
12	High School senior
11	junior
10	sophomore
9	freshman
8	Eight grade
7	Seventh grade
6	Sixth grade

ATICALL
Nelle Ott and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IR4LL (Ott 2009)
Measuring Text Difficulty
Readability Formulas
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary
Appendix
Prototype screenshots
IR4CALL NLP pipeline

ERHARDT-KARL
UNIVERSITÄT
TÜBINGEN

16 / 45

Traditional Readability Formulas

- ▶ Over two hundred traditional readability formulas have been developed (cf. Dubey 2004).
- ▶ They are generally developed for special purposes, such as determining the complexity of military training manuals (Caylor et al. 1973).
- ▶ A frequently used traditional readability measure is the Flesch-Kincaid formula (Kincaid et al. 1975)

ATICALL
Nils Oti and Damar Meurers

Introduction
Pedagogical grounding

The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty

Readability Formulae

Local Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype screenshots
IRCALL NLP pipeline

Example: Flesh-Kincaid

- ▶ Computes U.S. grade level needed to read a text.
- ▶ Derived empirically from set of hand-classified documents.

$$\text{Flesch-Kincaid} = -15.59 + 11.8 \cdot \text{AWL}_S + 0.39 \cdot \text{ASL}$$

Where

$$\text{AWL}_S = \frac{\text{Number of Syllables}}{\text{Number of Words}} \quad \text{Average word length counted in syllables.}$$

$$\text{ASL} = \frac{\text{Number of Words}}{\text{Number of Sentences}} \quad \text{Average sentence length.}$$

- ▶ Idea:
 - ▶ The longer the word, the harder it is. (and the less common it is, cf. Zipf 1936)
 - ▶ The longer the sentence, the harder it is to understand.

ATICALL
Nils Oti and Damar Meurers

Introduction
Pedagogical grounding

The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty

Readability Formulae

Local Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype screenshots
IRCALL NLP pipeline

Another example: Dale & Chall (1948)

$$\text{Dale-Chall} = 0.1579 \cdot \text{DS} + 0.0496 \cdot \text{ASL} + 3.6365$$

Where

DS = Dale Score The percentage of words outside the Dale list of 3000 words.

$$\text{ASL} = \frac{\text{Number of Words}}{\text{Number of Sentences}} \quad \text{Average sentence length.}$$

- ▶ Adds the idea of a specific list of "easy" words.
- ▶ List produced by "testing forth-graders on their knowledge in reading of a list of approximately 10,000 words".
- ▶ The more words are outside the set of "easy" words, the more difficult the text is.

ATICALL
Nils Oti and Damar Meurers

Introduction
Pedagogical grounding

The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty

Readability Formulae

Local Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype screenshots
IRCALL NLP pipeline

Traditional readability measures: Evaluation

- ▶ Pros:
 - ▶ Relatively simple to use.
 - ▶ 'Simple' NLP only: tokenizer, stemming, sentence splitter, sometimes syllable counter
- ▶ Cons:
 - ▶ Originally developed and validated using very small and often highly specific data sets (e.g., technical manuals).
 - ▶ No explicit validation of automatic analysis compared to original human analysis (e.g., syllable counting)
 - ▶ Measures such as sentence length are relative to domain.
 - ▶ Underlying assumptions (e.g., 'long sentences are difficult') are rather crude generalizations.

ATICALL
Nils Oti and Damar Meurers

Introduction
Pedagogical grounding

The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty

Readability Formulae

Local Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype screenshots
IRCALL NLP pipeline

Lexical Frequency Profiles (LFPs)

- ▶ Introduced by Laufer & Nation (1995) for the purpose of measuring the vocabulary used by learners.
- ▶ Ott (2009) uses LFPs 'upside down': measuring vocabulary in texts for learners, not by learners.
- ▶ LFPs work with 3 word lists:
 - ▶ First 1000 words of the General Service List (West 1953).
 - ▶ General Service List: list of words sorted by frequency
 - ▶ Second 1000 words of the General Service List.
 - ▶ Academic Word List (Coxhead 2000).
 - ▶ Underlying assumption: lists are mutually exclusive.

ATICALL
Nelle Ott and Damar Meurers

Introduction
Pedagogical grounding

The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulae

General Frequency Profiles

Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype screenshots
IRCALL NLP pipeline

Lexical Frequency Profile: Example

Results for a typical Wikipedia article:

Word List	Tokens		Types		Families
GSL 1	2202	75.39%	542	54.25%	384
GSL 2	121	4.14%	94	9.41%	78
AWL	245	8.39%	136	13.61%	109
Others	353	12.08%	227	22.72%	n.a.
Total	2921	100%	999	100%	n.a.

- ▶ Families: related by simple morphological processes
 - ▶ e.g., *happy*, *happily*, and *happyness* are in same family

ATICALL
Nelle Ott and Damar Meurers

Introduction
Pedagogical grounding

The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulae

General Frequency Profiles

Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype screenshots
IRCALL NLP pipeline

Vocabulary-based measures

- ▶ Pros:
 - ▶ Vocabulary is an important issue for learners.
 - ▶ 'Simple' NLP only: tokenizer, lemmatizer, perhaps tagger.
 - ▶ Measure can be informed by controlled vocabulary lists of text books.
 - ▶ Lists can also be extracted from corpora.
- ▶ Cons:
 - ▶ Vocabulary **changes** constantly, e.g., the General Service List was published in 1953 and correspondingly does not contain words such as *Internet* or *e-mail*?
 - ▶ Vocabulary is **domain-specific**:
Does the Academic Word List contain words of your field of research?

ATICALL
Nelle Ott and Damar Meurers

Introduction
Pedagogical grounding

The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulae

General Frequency Profiles

Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype screenshots
IRCALL NLP pipeline

Syntactic Complexity

- ▶ Vocabulary useful indicator, but if sentences are complex, learners will still have trouble understanding them.
- ▶ Sentence length as used in readability formulas simplistic.
- ▶ How can syntactic complexity be measured?
- ▶ Two simple units (Hunt 1965):
 - ▶ Clause: "a structure with a subject and a finite verb"
 - ▶ T-unit: "a main clause plus any subordinate clauses"

ATICALL
Nelle Ott and Damar Meurers

Introduction
Pedagogical grounding

The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulae

General Frequency Profiles

Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype screenshots
IRCALL NLP pipeline

Measuring syntactic complexity

Lu (2009) automates 14 measures of syntactic complexity which have been discussed as correlating with L2 proficiency:

Type	Measure
Length of production	Mean length of clause Mean length of sentence Mean length of T-unit
Sentence complexity	Mean number of clauses per sentence
Subordination	Mean number of clauses per T-unit Mean number of complex T-units per T-unit Mean number of dependent clauses per clause Mean number of dependent clauses per T-unit
Coordination	Mean number of coordinate phrases per clause Mean number of coordinate phrases per T-unit Mean number of T-units per sentence
Particular structures	Mean number of complex nominals per clause Mean number of complex nominals per T-unit Mean number of verb phrases per T-unit

ATICALL
Nelle Oll and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulae
Lexical Frequency Profiles
Syntactic Complexity

Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Tweaks Evaluation

Summary

Appendix
Prototype screenshots
IRCALL NLP pipeline

ERHARD-KARL
UNIVERSITÄT
TÜBINGEN

25 / 45

Textbook structures

- ▶ Textbooks introduce linguistic categories and forms in order of perceived complexity.
- ▶ For the purpose of teaching grammar, particular structures are especially relevant, e.g. 'give me a text with a lot of gerunds'.
 - ▶ Ott & Ziai (2008) developed a constraint grammar-based approach for classifying *-ing* forms into gerunds, participles, and the progressive forms.

ATICALL
Nelle Oll and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulae
Lexical Frequency Profiles
Syntactic Complexity

Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Tweaks Evaluation

Summary

Appendix
Prototype screenshots
IRCALL NLP pipeline

ERHARD-KARL
UNIVERSITÄT
TÜBINGEN

26 / 45

Textbook structures: Example

Linguistic structures taught in a textbook for English (Klett: Green Line 4, Weisshaar 2008):

Unit	Structures taught
1	Present perfect progressive with <i>since</i> and <i>for</i> Past perfect progressive Attributive use of adjectives after nouns Adverbs of degree
2	Perfect infinitive with modal verbs Passive infinitive with full verbs and modals
3	Gerund as subject, object, and after verbs and adjectives with prepositions Object plus <i>-ing</i> form Present and past progressive passive Passive with verbs with prepositions
4	Verb plus object plus infinitive Infinitive after question words and after superlatives Infinitives vs. Gerund
5	Non-defining relative clauses Participles as adjectives

ATICALL
Nelle Oll and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulae
Lexical Frequency Profiles
Syntactic Complexity

Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Tweaks Evaluation

Summary

Appendix
Prototype screenshots
IRCALL NLP pipeline

ERHARD-KARL
UNIVERSITÄT
TÜBINGEN

27 / 45

Information Retrieval

Manning et al. (2008, ch. 1):

"Information Retrieval is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers)."

ATICALL
Nelle Oll and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulae
Lexical Frequency Profiles
Syntactic Complexity

Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Tweaks Evaluation

Summary

Appendix
Prototype screenshots
IRCALL NLP pipeline

ERHARD-KARL
UNIVERSITÄT
TÜBINGEN

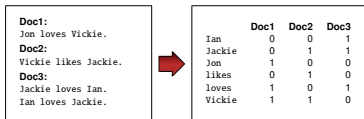
28 / 45

Indexing does the trick in IR!

Simply put:

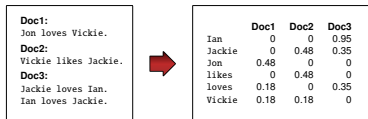
- ▶ Usually one has documents that contain words ("terms").
- ▶ Re-sort everything so that one has terms that are associated with documents → **indexing**.
- ▶ Result: the terms from the query can be mapped to terms in the index at low cost, giving you the corresponding documents quickly.

Example: Boolean index



Index with weights: Example

- ▶ TF-IDF (Term Frequency · Inverse Document Frequency):
Weigh terms which occur in fewer documents more highly.



Text models

- ▶ In addition to the words themselves, any information about a text can be used as an index.
 - ▶ Here: readability measures
- ▶ All measures are stored in a table for each text, the so-called text model.
- ▶ The table contains the key (name) for each measure and a value.

Example of a text model (extract)

Type	Key	Value
General	Character Count	14249
General	Sentence Count	111
General	Token Count	2542
General	Type-Token Ratio	0.3703
LFP	Academic Word List Token Ratio	0.0816
LFP	Academic Word List Type Ratio	0.1389
LFP	General Service List 1k Token Ratio	0.1389
LFP	General Service List 1k Type Ratio	0.4191
LFP	General Service List 2k Token Ratio	0.0557
LFP	General Service List 2k Type Ratio	0.0841
LFP	Off-List Token Ratio	1.3119
LFP	Off-List Type Ratio	0.1325
Readability	Automatic Readability Index	12.7182
Readability	Flesch Reading Ease	57.6363
Readability	Gunning Fog Index	19.4510
Readability	Original Dale-Chall Score	8.8971

Demo: <http://drni.de/zap/ir4ll>

INFORMATION RETRIEVAL FOR LANGUAGE LEARNING
A Search Engine Prototype

Search:

Enter your query and select your reading level
(Group) [easy] [medium] [hard]

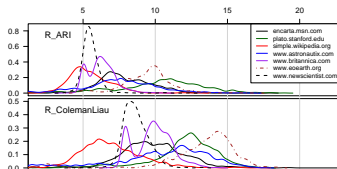
- [1. Read: drugs - Simple English Wikipedia, the free encyclopedia](#)
Read: drugs can be made from three. [Quanta on Read: drugs](#) in many countries. [Read: drugs - drugs](#) that come in powder form. [Read: drugs](#) are often "used" the [Read: drugs](#) that their use, makes it a crime. [Read: drugs](#) in effect to [Read: drugs](#) into their body by trying to eat water. Some [Read: drugs](#) can be cracked. [Read: drugs](#) from Wikipedia, the free encyclopedia, and/or navigation, search these are easy.
- [2. Drugs - Simple English Wikipedia, the free encyclopedia](#)
Drugs From Wikipedia, the free encyclopedia Jump to: navigation, search [Drugs](#) may mean ... or recreational [drugs](#) - Chemical salts, liquids, or powders of plants that people take to make them feel ... page Retrieved from <http://simple.wikipedia.org/w/index.php?title=Drugs&oldid=104849> Category: Information News Page
- [3. Brain-boosting drugs: not to be feared - health - 14 December 2008 - New Scientist](#)
A different type of medication with the name [drugs](#) and the old [drugs](#) they certainly are not. [14 Dec](#) [Brain-boosting drugs](#) not to be feared 14 December 2008 Register Free 2008. Subscribe and get 4 free issues. For order stories, visit the [Home](#) [Brain and Drugs](#) and Alcohol Topic. [14 Dec](#) [Brain-boosting drugs](#) not to be feared 14 December 2008 Register Free 2008. Subscribe
- [4. Quotes: Alcohol and Drugs - health - 01 September 2009 - New Scientist](#)
[14 Dec](#) [Brain-boosting drugs](#) not to be feared 14 December 2008 by New Scientist For order stories, visit the [Drugs](#) and Alcohol Topic. Quotes: May not be as hard on the brain before the discoveries ... in the US new research shows the influence of alcohol on [drugs](#) in 1997 - 10 September. [14 Dec](#) [Brain-boosting drugs](#) not to be feared 14 December 2008 by New Scientist For order stories

Towards Evaluation

- ▶ An experiment with 190.872 unique documents downloaded from 7 online encyclopedias.
- ▶ Encyclopedias are likely to contain articles on one topic each, but with different text difficulty.
- ▶ Sample of 7.000 text models (1.000 models for each site).

Towards Evaluation: Some results

Distribution of scores from two grade level-based measures:



- ▶ This type of evaluation gives only a first impression.
- ▶ A gold standard (annotated corpus) should be created and used instead.

ATICALL
Nelsi Ott and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulas
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation
Summary
Appendix
Prototype screenshots
IRCALL NLP pipeline

EMERSON COLLEGE
UNIVERSITÄT
TÜBINGEN
33 / 45



ATICALL
Nelsi Ott and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulas
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation
Summary
Appendix
Prototype screenshots
IRCALL NLP pipeline

EMERSON COLLEGE
UNIVERSITÄT
TÜBINGEN
35 / 45



ATICALL
Nelsi Ott and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulas
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation
Summary
Appendix
Prototype screenshots
IRCALL NLP pipeline

EMERSON COLLEGE
UNIVERSITÄT
TÜBINGEN
34 / 45



ATICALL
Nelsi Ott and Damar Meurers

Introduction
Pedagogical grounding
The WERTI system
Modeling FLT practice
Progression in WERTI
Examples and Target types
Realizing proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulas
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation
Summary
Appendix
Prototype screenshots
IRCALL NLP pipeline

EMERSON COLLEGE
UNIVERSITÄT
TÜBINGEN
36 / 45



Summary

- ▶ Fostering language awareness is a well-motivated component of FLT.
- ▶ We discussed WERTi: web-based activity generator based on real-world texts selected by the learner.
 - ▶ a learner-driven approach, in which learners can
 - ▶ generate as many activities as they want
 - ▶ choose texts that match their interests
 - ▶ activities that remain fully contextualized as whole articles with the original web presentation intact
 - ▶ learner interaction with simple feedback based on the original text and linguistic analysis
- ▶ Develop search for real-world texts supporting a range of reading difficulty measures and specific linguistic categories → IR4LL.

ATICALL

Niels Ott and Detmar Meurers

Introduction

Pedagogical grounding

The WERTi system

Modeling FLT practice

Progression in WERTi

Examples and Target types

Realizing proposal

Creating ex. progression

Prototype realization

Some challenges

IR4LL (Ott 2009)

Measuring Text Difficulty

Readability Formulas

Lexical Frequency Profiles

Syntactic Complexity

Textbook structures

A Search Engine Prototype

Information Retrieval

Adding Text Difficulty

Towards Evaluation

Summary

Appendix

Prototype screenshots

IR4CALL NLP pipeline



37/45

References

- Amaral, L., V. Metcalf & D. Meurers (2006). Language Awareness through Re-use of NLP Technology. Pre-conference Workshop on NLP in CALL – Computational and Linguistic Challenges. CALICO 2006. May 17, 2006. University of Hawaii.
- Bird, S. & E. Loper (2004). NLTK: The Natural Language Toolkit. In *Proceedings of the ACL demonstration session*. Barcelona, Spain: Association for Computational Linguistics, pp. 214–217. URL <http://aclweb.org/anthology/P04-3031>.
- Caylor, J. S., T. G. Sticht, L. C. Fox & J. P. Ford. (1973). *Methodologies for determining reading requirements of military occupational specialties: Technical report No. 73-5*. Tech. rep., Human Resources Research Organization, Alexandria, VA.
- Coxhead, A. (2000). A New Academic Word List. *Teachers of English to Speakers of Other Languages* 34(2), 213–238.
- Dale, E. & J. S. Chall (1948). A Formula for Predicting Readability. *Educational research bulletin; organ of the College of Education* 27(1), 11–28.
- Dickinson, M. (2005). Error detection and correction in annotated corpora. Ph.D. thesis, The Ohio State University.
- Dickinson, M. & W. D. Meurers (2003). Detecting Errors in Part-of-Speech Annotation. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL-03)*. Budapest, Hungary, pp. 107–114. URL <http://purl.org/dm/papers/dickinson-meurers-03.html>.

ATICALL

Niels Ott and Detmar Meurers

Introduction

Pedagogical grounding

The WERTi system

Modeling FLT practice

Progression in WERTi

Examples and Target types

Realizing proposal

Creating ex. progression

Prototype realization

Some challenges

IR4LL (Ott 2009)

Measuring Text Difficulty

Readability Formulas

Lexical Frequency Profiles

Syntactic Complexity

Textbook structures

A Search Engine Prototype

Information Retrieval

Adding Text Difficulty

Towards Evaluation

Summary

Appendix

Prototype screenshots

IR4CALL NLP pipeline



37/45

- Dubey, A. (2004). Statistical Parsing for German: Modeling Syntactic Properties and Annotation Differences. Ph.D. thesis, Universität des Saarlandes.
- Ellis, N. (1994). Implicit and Explicit Language Learning - An Overview. In *Implicit and Explicit Learning of Languages*, San Diego, CA: Academic Press, pp. 1–31.
- Gunning, R. (1968). *The Technique of Clear Writing*. New York: McGraw-Hill Book Company, 2nd ed.
- Hunt, K. W. (1965). Grammatical Structures Written at Three Grade Levels. NCTE Research Report No. 3.
- Kincaid, J. P., R. P. J. Fishburne, R. L. Rogers & B. S. Chissom (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel*. Research Branch Report 8-75, Naval Technical Training Command, Millington, TN.
- Klare, G. R. (1963). *The Measurement of Readability*. Ames, Iowa: Iowa State University Press.
- Lauler, B. & P. Nation (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics* 16(3), 307–322.
- Lightbown, P. M. & N. Spada (1999). *How languages are learned*. Oxford: Oxford University Press.
- Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. D. Bot, C. Kramsch & R. Ginsberg (eds.), *Foreign language research in cross-cultural perspective*, Amsterdam: John Benjamins, pp. 39–52.
- Long, M. H. (1996). The role of linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (eds.), *Handbook of second language acquisition*, New York: Cambridge University Press, pp. 413–468.

ATICALL

Niels Ott and Detmar Meurers

Introduction

Pedagogical grounding

The WERTi system

Modeling FLT practice

Progression in WERTi

Examples and Target types

Realizing proposal

Creating ex. progression

Prototype realization

Some challenges

IR4LL (Ott 2009)

Measuring Text Difficulty

Readability Formulas

Lexical Frequency Profiles

Syntactic Complexity

Textbook structures

A Search Engine Prototype

Information Retrieval

Adding Text Difficulty

Towards Evaluation

Summary

Appendix

Prototype screenshots

IR4CALL NLP pipeline



37/45

- Lu, X. (2009). Automatic measurement of syntactic complexity in child language acquisition. *International Journal of Corpus Linguistics* 14, 3–28(26). URL <http://www.ingentaconnect.com/content/ijcl/2009/00000014/00000001/art00002>.
- Lyster, R. (1998). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classroom. *Language Learning* 48, 183–218.
- Manning, C. D., P. Raghavan & H. Schütze (2008). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Metcalf, V. & D. Meurers (2006). Generating Web-based English Preposition Exercises from Real-World Texts. URL <http://www.ling.ohio-state.edu/call/handouts/eurocall06-metcalf-meurers.pdf>. EUROCALL 2006. Granada, Spain. September 4–7, 2006.
- Norris, J. & L. Ortega (2000). Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-Analysis. *Language Learning* 50(3), 415–528.
- Ott, N. (2009). Information Retrieval for Language Learning: An Exploration of Text Difficulty Measures. Master's thesis, University of Tübingen, Seminar für Sprachwissenschaft, Tübingen, Germany. URL <http://drni.de/zip/ma-thesis>.
- Ott, N. & R. Ziai (2008). ICALL Activities for Gerunds vs. To-infinitives: A Constraint-Grammar-based extension to the New WERTi System. URL <http://www.drni.de/niels/n3files/read-me/werti-gerunds.pdf>. Unpublished term paper for the course *Using Natural Language Processing to Foster Language Awareness in Second Language Learning* taught in summer 2008 at Tübingen University by Detmar Meurers.

ATICALL

Niels Ott and Detmar Meurers

Introduction

Pedagogical grounding

The WERTi system

Modeling FLT practice

Progression in WERTi

Examples and Target types

Realizing proposal

Creating ex. progression

Prototype realization

Some challenges

IR4LL (Ott 2009)

Measuring Text Difficulty

Readability Formulas

Lexical Frequency Profiles

Syntactic Complexity

Textbook structures

A Search Engine Prototype

Information Retrieval

Adding Text Difficulty

Towards Evaluation

Summary

Appendix

Prototype screenshots

IR4CALL NLP pipeline



37/45

Schmid, H. (1994). Part-of-Speech Tagging Using Decision Trees. In *Proceedings of the International Conference on New Methods in Language Processing*. Manchester, UK. URL <http://www.ims.uni-stuttgart.de/fp/pub/corpora/tree-tagger1.pdf>.

Schmidt, R. (1995). Consciousness and foreign language: A tutorial on the role of attention and awareness in learning. In R. Schmidt (ed.), *Attention and awareness in foreign language learning*. Honolulu: University of Hawaii Press, pp. 1–63.

Schulz, R. A. (2002). Hilft es die Regel zu wissen um sie anzuwenden? Das Verhältnis von metalinguistischem Bewusstsein und grammatischer Kompetenz in DaF. *Die Unterrichtspraxis—Teaching German* 35(1), 15–24. URL <http://www.jstor.org/stable/pdfplus/3531951.pdf>.

Weisshaar, H. (2008). *Green Line Schülerbuch 4 – Band 4: 8. Klasse*. Stuttgart, Germany: Ernst Klett.

West, M. (1953). *A General Service List of English Words*. London: Longmans.

Zipf, G. K. (1936). *The Psycho-Biology of Language*. London: Routledge.

ATICALL
Nelle Ot and Damar Meurers

Introduction
Pedagogical grounding
The WERTi system
Modeling FLT practice
Progression in WERTi
Examples and Target types
Reading proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulas
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype schemata
IRCALL NLP pipeline

ERHARD-KARL UNIVERSITÄT TüBINGEN
38 / 45

Working With English Real Texts:
An Intelligent Workbook for English

Login
Username:
Password:
Submit

The WERTi Project
About
People
Acknowledgements

OSU Project Context
CLT research group
Dept. of Linguistics
Humanities Inf. Syst.

WERTi
WORKING WITH ENGLISH REAL-TEXTS:
AN INTELLIGENT WORKBOOK FOR ENGLISH

ERHARD-KARL UNIVERSITÄT TüBINGEN
38 / 45

ATICALL
Nelle Ot and Damar Meurers

Introduction
Pedagogical grounding
The WERTi system
Modeling FLT practice
Progression in WERTi
Examples and Target types
Reading proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulas
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype schemata
IRCALL NLP pipeline

ERHARD-KARL UNIVERSITÄT TüBINGEN
38 / 45

Working With English Real Texts:
An Intelligent Workbook for English

WERTi workbook topics
Prepositions
Pronouns
Active/Passive

The WERTi Project
About
People
Acknowledgements

OSU Project Context
CLT research group
Dept. of Linguistics
Humanities Inf. Syst.

Welcome to WERTi!

What is an "intelligent workbook"?

WERTi is a "workbook" because it provides you with activities for a number of grammar topics, and "intelligent" because it makes up those activities when you ask it to, using articles you choose yourself!

How do I use WERTi?

1. First, choose a **workbook topic** from the list on the left, or at the bottom of this page.
2. Once you have chosen a workbook topic, WERTi will ask you to enter a **search topic** you are interested in. It will find articles on that topic. You choose whichever one you like.
3. Finally, choose an **activity**. Activity types are explained next.

What activities can I choose from?

- **Color**
WERTi will find all the examples of your workbook topic in the article and color them blue. If you are interested in prepositions, for example, WERTi will show you all the prepositions in the article.
- **Click**
This time it is your turn to find examples of your workbook topic in the article, and then click on them. If you are looking for prepositions, and you click on one, it will turn green. If you click on something else, it will turn red.
- **Practice**
WERTi will provide you with one or more activities to let you practice using examples of your workbook topic. WERTi will ask you to fill in blanks, or rearrange words by clicking on them and dragging them, or find and fix mistakes.

What workbook topics can I choose from?

- Prepositions
- Pronouns
- Active/Passive

ERHARD-KARL UNIVERSITÄT TüBINGEN
39 / 45

ATICALL
Nelle Ot and Damar Meurers

Introduction
Pedagogical grounding
The WERTi system
Modeling FLT practice
Progression in WERTi
Examples and Target types
Reading proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulas
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype schemata
IRCALL NLP pipeline

ERHARD-KARL UNIVERSITÄT TüBINGEN
39 / 45

Working With English Real Texts:
An Intelligent Workbook for English

Prepositions

Search in [Reuters news](#) to obtain a text you want to work with:

ROWS

[Reuters News Search](#)

Preposition Activities

- **Color**
WERTi shows you all the prepositions in the text in blue.
- **Click**
Find the prepositions and click on them.
- **Fill-in-the-blanks**
Fill in all the prepositions in the article.

ERHARD-KARL UNIVERSITÄT TüBINGEN
40 / 45

ATICALL
Nelle Ot and Damar Meurers

Introduction
Pedagogical grounding
The WERTi system
Modeling FLT practice
Progression in WERTi
Examples and Target types
Reading proposal
Creating ex. progression
Prototype realization
Some challenges

IRALL (Ott 2009)
Measuring Text Difficulty
Readability Formulas
Lexical Frequency Profiles
Syntactic Complexity
Textbook structures
A Search Engine Prototype
Information Retrieval
Adding Text Difficulty
Towards Evaluation

Summary

Appendix
Prototype schemata
IRCALL NLP pipeline

ERHARD-KARL UNIVERSITÄT TüBINGEN
40 / 45

WERTI *Working With English Real Texts:*
An Intelligent Workbook for English

Search Home Logout Help

Prepositions

Select an activity based on the results of your search:

Moove slowly and don't hug cows, hikers told (Wednesday, August 30, 2006 1:01 AM ET)
Evelyn Zaugg of the Swiss Hiking Federation said that while there were no precise statistics on incidents involving cows, walkers are reporting more run-ins...

Cows 'moo' with an accent, farmers believe (Thursday, August 24, 2006 2:04 AM ET)
LONDON (Reuters) - Cows have regional accents, a group of British farmers claims, and phonetics experts say the idea is not as far-fetched as it sounds...

US drivers subsidize European pump prices - report (Thursday, August 21, 2006 9:13 AM ET)
America consumers have become the "cash cows" for the international oil industry, the study said, unlike U.S...

Accidental death of bear fuels passions in France (Wednesday, August 30, 2006 5:53 AM ET)
spokesman for ASPAP, a group that represents farmers who have protested against the reintroduction program which they say threatens sheep and cows that graze...

41 / 45

ATICALL

Nais Cit and Demar Mueurs

Introduction

Pedagogical grounding

The WERTI system

Modeling FLT practice

Examples and Target type

Realizing proposal

Creating ex. progression

Prototype realization

Some challenges

IRALL (Oct 2009)

Measuring Test Difficulty

Reliability Formulas

Lexical Frequency Profiles

Sprynck Complexity

Textbook structures

A Search Engine Prototype

Information Retrieval

Adding Test Difficulty

Towards Evaluation

Summary

Prototype assessments

IRCALL NLP pipeline

Appendix

Prototype assessments

IRCALL NLP pipeline



REUTERS

You are here: Home > News > Daily > English > Article

Go to Section: US International Business Markets Public Entertainment Technology Sports Daily Enough US-Open Coverage

Cows 'moo' with an accent, farmers believe

The Aug 24, 2006 2:04 AM ET

End This Article Print This Article Reports

LONDON (Reuters) - Cows have regional accents, a group of British farmers claims, and phonetics experts say the idea is not as far-fetched as it sounds.

Ulyss Green, from southwest England, was one of a group of farmers who first noticed the phenomenon.

"I spent a lot of time with my Frisians and they definitely 'moo' with a Somerset drawl," he said, "referring to the breed of dairy cow he owns."

"I've spoken to the other farmers in the West Country group and they have noticed a similar development in their own herds."

"I think it works the same with dogs - the closer a farmer's bond is with his animals, the easier it is for them to pick up his accent."

Don Lane, spokesman for a group called the West Country Farmhouse Cheesemakers to which Green belongs, said a contracted John Wells, Professor of Phonetics at University College London, who said that a similar phenomenon had been found in birds.

"You find distinct clipping accents in the same species around the country. This could also be true of cows," Wells said on the group's Web site (expedition.co.uk).

According to Lane, accents among cows probably develop in a similar way as among humans, and resulted from spending time with farmers with differing accents.

"Apparently the biggest influence on accents is peer-group or children in the playground, for example," he said. "Herds are quite tight-knit communities and do tend to learn the area."

He added that more scientific research was needed to prove what was just an anecdotal theory at this stage.

© Reuters 2006. All Rights Reserved.

REUTERS

You are here: Home > News > Daily > English > Article

Go to Section: US International Business Markets Public Entertainment Technology Sports Daily Enough US-Open Coverage

Cows 'moo' with an accent, farmers believe

The Aug 24, 2006 2:04 AM ET

End This Article Print This Article Reports

LONDON (Reuters) - Cows have regional accents, a group of British farmers claims, and phonetics experts say the idea is not as far-fetched as it sounds.

Ulyss Green, from southwest England, was one of a group of farmers who first noticed the phenomenon.

"I spent a lot of time with my Frisians and they definitely 'moo' with a Somerset drawl," he said, "referring to the breed of dairy cow he owns."

"I've spoken to the other farmers in the West Country group and they have noticed a similar development in their own herds."

"I think it works the same with dogs - the closer a farmer's bond is with his animals, the easier it is for them to pick up his accent."

Don Lane, spokesman for a group called the West Country Farmhouse Cheesemakers to which Green belongs, said a contracted John Wells, Professor of Phonetics at University College London, who said that a similar phenomenon had been found in birds.

"You find distinct clipping accents in the same species around the country. This could also be true of cows," Wells said on the group's Web site (expedition.co.uk).

According to Lane, accents among cows probably develop in a similar way as among humans, and resulted from spending time with farmers with differing accents.

"Apparently the biggest influence on accents is peer-group or children in the playground, for example," he said. "Herds are quite tight-knit communities and do tend to learn the area."

He added that more scientific research was needed to prove what was just an anecdotal theory at this stage.

© Reuters 2006. All Rights Reserved.



ATICALL

Nais Cit and Demar Mueurs

Introduction

Pedagogical grounding

The WERTI system

Modeling FLT practice

Examples and Target type

Realizing proposal

Creating ex. progression

Prototype realization

Some challenges

IRALL (Oct 2009)

Measuring Test Difficulty

Reliability Formulas

Lexical Frequency Profiles

Sprynck Complexity

Textbook structures

A Search Engine Prototype

Information Retrieval

Adding Test Difficulty

Towards Evaluation

Summary

Prototype assessments

IRCALL NLP pipeline

Appendix

Prototype assessments

IRCALL NLP pipeline



REUTERS

You are here: Home > News > Daily > English > Article

Go to Section: US International Business Markets Public Entertainment Technology Sports Daily Enough US-Open Coverage

Cows 'moo' with an accent, farmers believe

The Aug 24, 2006 2:04 AM ET

End This Article Print This Article Reports

LONDON (Reuters) - Cows have regional accents, a group of British farmers claims, and phonetics experts say the idea is not as far-fetched as it sounds.

Ulyss Green, from southwest England, was one of a group of farmers who first noticed the phenomenon.

"I spent a lot of time with my Frisians and they definitely 'moo' with a Somerset drawl," he said, "referring to the breed of dairy cow he owns."

"I've spoken to the other farmers in the West Country group and they have noticed a similar development in their own herds."

"I think it works the same with dogs - the closer a farmer's bond is with his animals, the easier it is for them to pick up his accent."

Don Lane, spokesman for a group called the West Country Farmhouse Cheesemakers to which Green belongs, said a contracted John Wells, Professor of Phonetics at University College London, who said that a similar phenomenon had been found in birds.

"You find distinct clipping accents in the same species around the country. This could also be true of cows," Wells said on the group's Web site (expedition.co.uk).

According to Lane, accents among cows probably develop in a similar way as among humans, and resulted from spending time with farmers with differing accents.

"Apparently the biggest influence on accents is peer-group or children in the playground, for example," he said. "Herds are quite tight-knit communities and do tend to learn the area."

He added that more scientific research was needed to prove what was just an anecdotal theory at this stage.

© Reuters 2006. All Rights Reserved.



ATICALL

Nais Cit and Demar Mueurs

Introduction

Pedagogical grounding

The WERTI system

Modeling FLT practice

Examples and Target type

Realizing proposal

Creating ex. progression

Prototype realization

Some challenges

IRALL (Oct 2009)

Measuring Test Difficulty

Reliability Formulas

Lexical Frequency Profiles

Sprynck Complexity

Textbook structures

A Search Engine Prototype

Information Retrieval

Adding Test Difficulty

Towards Evaluation

Summary

Prototype assessments

IRCALL NLP pipeline

Appendix

Prototype assessments

IRCALL NLP pipeline



ATICALL

Nais Cit and Demar Mueurs

Introduction

Pedagogical grounding

The WERTI system

Modeling FLT practice

Examples and Target type

Realizing proposal

Creating ex. progression

Prototype realization

Some challenges

IRALL (Oct 2009)

Measuring Test Difficulty

Reliability Formulas

Lexical Frequency Profiles

Sprynck Complexity

Textbook structures

A Search Engine Prototype

Information Retrieval

Adding Test Difficulty

Towards Evaluation

Summary

Prototype assessments

IRCALL NLP pipeline

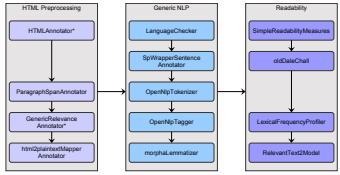
Appendix

Prototype assessments

IRCALL NLP pipeline



NLP pipelines in the indexer



Introduction

Pedagogical grounding

The WERTI system

Modeling FLT practice

Progression in WERTI

Examples and target types

Realizing proposal

Creating ex. progression

Prototype realization

Some challenges

IR4LL (Ott 2009)

Measuring Text Difficulty

Readability Formulas

Lexical Frequency Profiles

Syntactic Complexity

Textbook structures

A Search Engine Prototype

Information Retrieval

Adding Text Difficulty

Search Evaluation

Summary

Appendix

Prototype screenshots

ERICALL NLP pipeline