How Do Educational Settings at the Secondary Level Impact on Learners' Use of the English Passive? – Evidence from the Secondary-Level Corpus of Learner English (SCooLE)

Learner Corpus Research, Bergen/Norway, 27-29 September 2013

Verena Möller

Université catholique de Louvain
Centre for English Corpus Linguistics

Universität Hildesheim
Institut für Informationswissenschaft
und Sprachtechnologie

Overview

1. Educational Settings at the Secondary Level

2. The Passive as a Diagnostic Criterion

3. Compilation of the Corpora

   - TeaMC (Teaching Materials Corpus)

   - SCooLE (Secondary-Level Corpus of Learner English)

4. Evidence from the Corpora

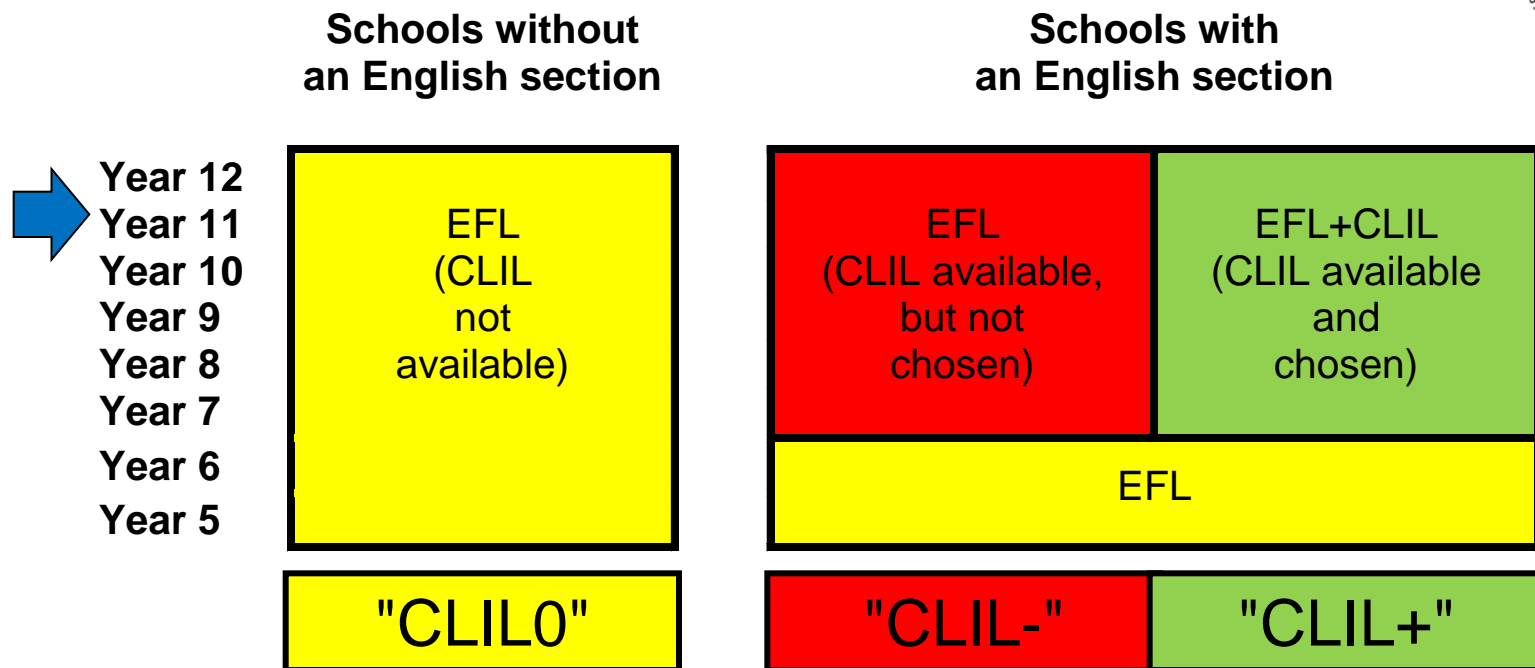   - Procedure

   - TeaMC

   - SCooLE

Educational Settings at the Secondary Level

Baden-Württemberg

Educational Settings at the Secondary Level

EFL and Content and Language Integrated Learning (CLIL)

	Schools without an English section	Schools with an English section	
Year 12 Year 11 Year 10 Year 9 Year 8 Year 7 Year 6 Year 5	EFL (CLIL not available)	EFL (CLIL available, but not chosen)	EFL+CLIL (CLIL available and chosen)
		EFL	
	"CLIL0"	"CLIL-"	"CLIL+"

BUT:

"Is CLIL so beneficial, or just selective?" (Bruton 2011)

The Passive as a Diagnostic Criterion

Motivation

Input:

CLIL materials differ from EFL materials in that they are scientifically oriented, i. e. they resemble scientific text

⇨ Research suggests that the passive is characteristic of scientific text

Svartvik 1966	*(scientific text):*	*19.3 pass./1,000 w.*
Wanner 2009	*(research abstracts):*	*17.0 pass./1,000 w.*
Wanner 2009	*(research abstracts):*	*25.2 % of VPs*
Holtz 2011	*(research abstracts):*	*55.7 % of VPs*
Holtz 2011	*(research articles):*	*46.6 % of VPs*

The Passive as a Diagnostic Criterion

Motivation

Strategies:

Less advanced learners may use synonymous active structures

Lexis-grammar interface:

Less advanced learners may prefer passives introduced by EFL materials as lexical chunks before the passive is introduced

Compilation of the TeaMC

Subcorpora/Linguistic annotation

TeaMC
(980,773 words)

TeaMC (input)	TeaMC (reference)
(568,328 words)	(412,445 words)

Year

7

8

9

10

11

12

EFL	CLIL
(149,015 words)	(419,313 words)
	• Geography
	(147,837 words)
	• History
	(202,596 words)
	• Politics
	(6,730 words)
	• Biology
	(62,150 words)

EFL
(412,445 words)

POS-Tagging: • TreeTagger (cf. Schmid 1994)
• CLAWS (cf. Garside/Smith 1997)

Compilation of the SCooLE

Text data

Discuss **TWO** of the following statements – choose **one from set I** and **one from set II**.

I.
1. In Germany, the education system offers equality of opportunity to everyone, rich or poor.
2. Minority groups should make greater efforts to integrate into the mainstream population.
3. Germany and the USA have a special relationship.
4. Privacy is a thing of the past.

II.
5. A better understanding between cultures can be created by travelling to other countries as a tourist.
6. The death penalty should be reintroduced in Germany.
7. In order to fight teenage drinking, the legal drinking age should be raised to 21.
8. In modern society, men and women are given equal chances.

Text 1:
⇨ Arbeitsplatz
⇨ Wechseldatenträger
⇨ *text1.txt*
⇨ rechte Maustaste "Öffnen mit"
⇨ Editor
⇨ Bitte das Speichern nicht vergessen!

Text 2:
⇨ Arbeitsplatz
⇨ Wechseldatenträger
⇨ *text2.txt*
⇨ rechte Maustaste "Öffnen mit"
⇨ Editor
⇨ Bitte das Speichern nicht vergessen!

Bitte arbeiten Sie NICHT mit *Word*.

Compilation of the SCooLE

Metadata on learner variables

STUDIE ZUR ENGLISCHEN LERNERSPRACHE BADEN-WÜRTTEMBERGISCHER GYMNASIASTEN

I. Persönliche Angaben:
1. Teilnehmernummer: _____
2. Alter: _____ Jahre
3. Geschlecht: ❑ männlich ❑ weiblich
4. Muttersprache (Sprache, in der zuerst das Sprechen gelernt wurde): (bei mehrsprachiger Erziehung seit der Geburt bitte alle betreffenden Sprachen angeben)
5. Sprachen, die täglich zu Hause gesprochen werden: (bitte alle betreffenden Sprachen angeben)
6. Gesamtdauer der Aufenthalte in englischsprachigen Ländern: ❑ keine ❑ kürzer als 3 Monate ❑ 3-6 Monate ❑ 6-12 Monate ❑ länger als 12 Monate (_____ Jahre)

II. Fremdsprachen:
7. Erste Fremdsprache: _____ seit ____ Jahren
   ❑ sehr gut ❑ eher gut ❑ mittel-mäßig ❑ eher schlecht ❑ sehr schlecht
8. Zweite Fremdsprache: _____ seit ____ Jahren
   ❑ sehr gut ❑ eher gut ❑ mittel-mäßig ❑ eher schlecht ❑ sehr schlecht
9. Dritte Fremdsprache: _____ seit ____ Jahren
   ❑ sehr gut ❑ eher gut ❑ mittel-mäßig ❑ eher schlecht ❑ sehr schlecht
10. Vierte Fremdsprache _____ seit ____ Jahren
    ❑ sehr gut ❑ eher gut ❑ mittel-mäßig ❑ eher schlecht ❑ sehr schlecht
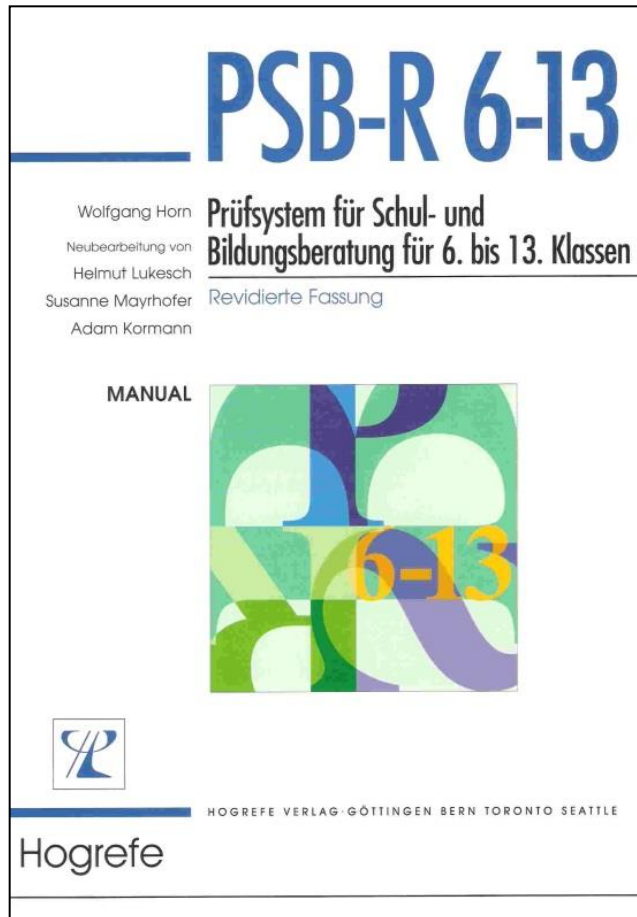11. Weitere Fremdsprachen: _____

Informal questionnaire:
- age
- gender
- L1
- language(s) spoken at home
- cumulative duration of stays in English-speaking countries
- other L2
- self-rated L2 competence
- school career
- educational settings attended
- spare time activities related to the English language
- etc.

Compilation of the SCooLE
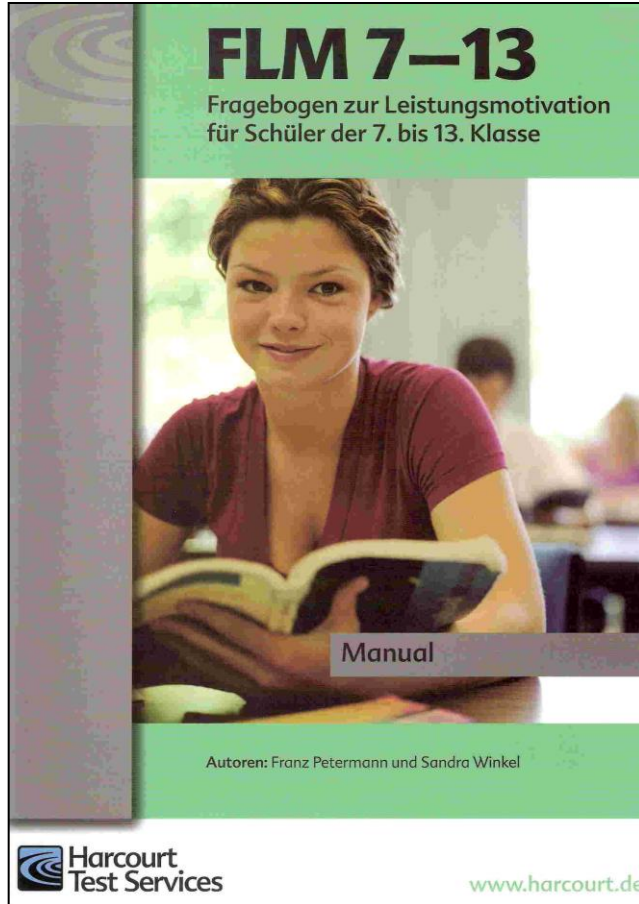
Metadata on learner variables

Psychometric test:

Aspects of intelligence

- overall
  - verbal
    - word recognition
    - word fluency
    - verbal reasoning
  - non-verbal reasoning
  - concentration
- etc.

Compilation of the SCooLE
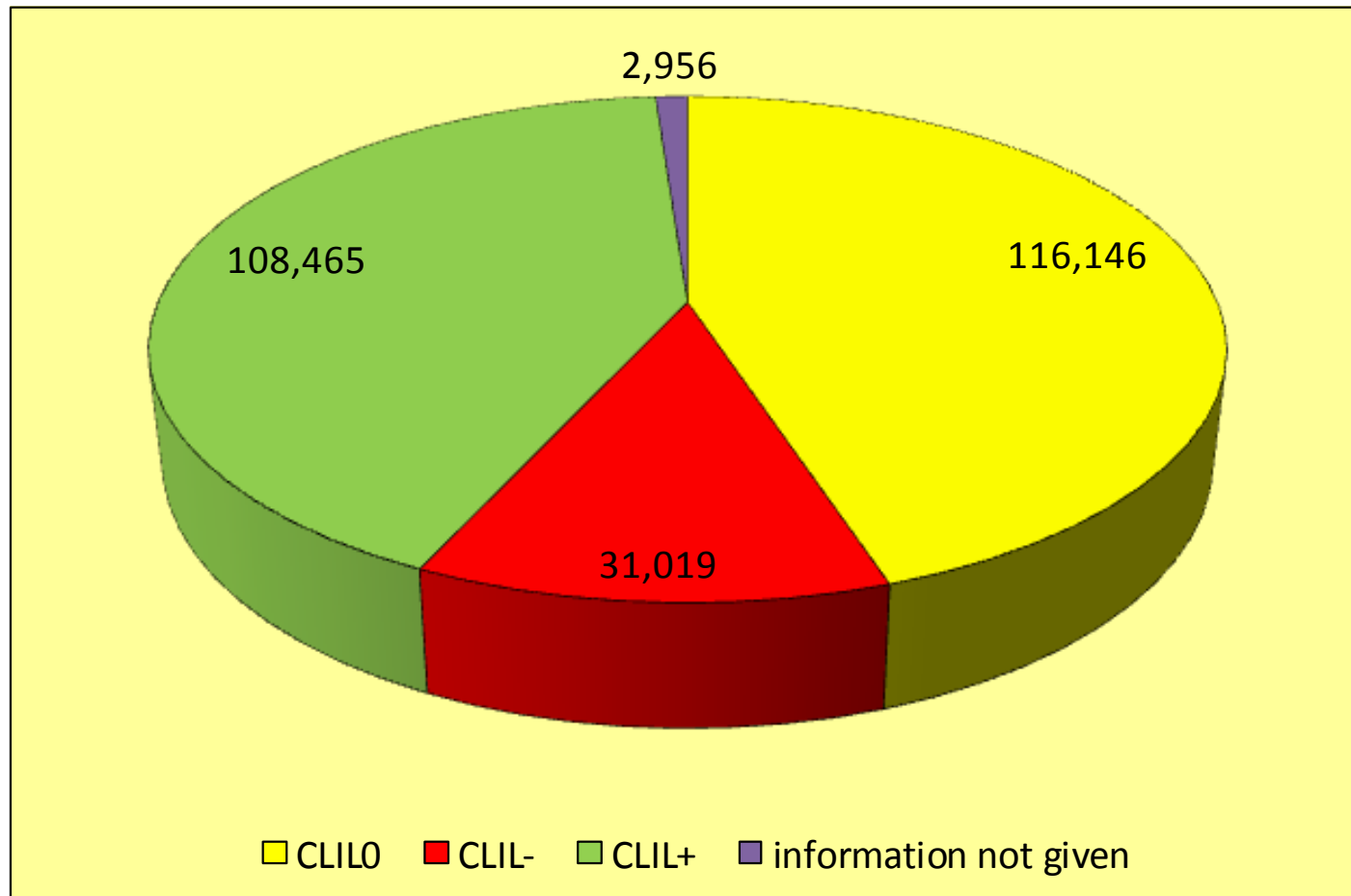
Metadata on learner variables

FLM 7–13
Fragebogen zur Leistungsmotivation
für Schüler der 7. bis 13. Klasse

Manual

Autoren: Franz Petermann und Sandra Winkel

Harcourt
Test Services

www.harcourt.de

Psychometric test:

Aspects of motivation

- orientation towards performance and success
- perseverance and effort
- etc.

Compilation of the SCooLE

Subcorpora

851 essays: > 250,000 words
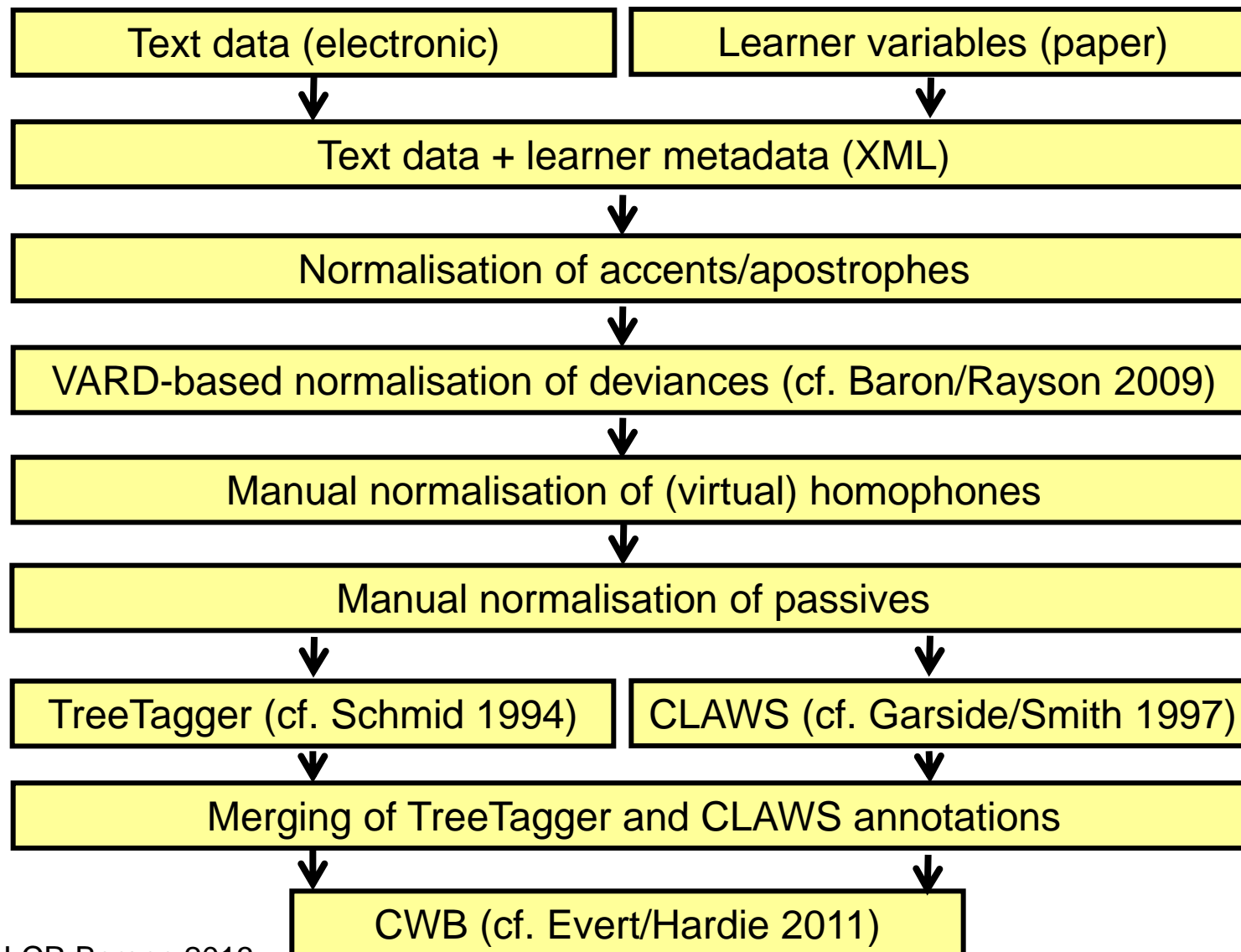
Compilation of the SCooLE

Challenges

Types of deviance:

- Omission of *be:*

    e. g.   *Should the death penalty reintroduced in Germany?*

- Morphological and/or orthographic errors in the form of *be* or related clitics:

    e. g.   *You arent forced to post anything in the internet.*

- Morphological and/or orthographic errors in the past participle:

    e. g.   *[…] alcohol can just be buyed by 21 old people.*

- Lexical errors:

    e. g.   *[…] so he is already prisoned by the police.*

- Overpassivization:

    e. g.   *[…] the murderer can not try to murder another human after they are released him from the prisoner.*

- etc.

Compilation of the SCooLE

Annotation

Text data (electronic)	Learner variables (paper)

↓ ↓

Text data + learner metadata (XML)

↓

Normalisation of accents/apostrophes

↓

VARD-based normalisation of deviances (cf. Baron/Rayson 2009)

↓

Manual normalisation of (virtual) homophones

↓

Manual normalisation of passives

↓ ↓

TreeTagger (cf. Schmid 1994)	CLAWS (cf. Garside/Smith 1997)

↓ ↓

Merging of TreeTagger and CLAWS annotations

↓ ↓

CWB (cf. Evert/Hardie 2011)

Compilation of the SCooLE

Recall rates for *beVed* (pilot study)

Procedure	TreeTagger	CLAWS
Initial	134/149 (89.9 %)	137/149 (91.9 %)
Normalisation of apostrophes	134/149 (89.9 %)	139/149 (93.3 %)
VARD-based normalisation	140/149 (94.0 %)	139/149 (93.3 %)
Manual normalisation of homophones	141/149 (94.6 %)	140/149 (94.0 %)
Manual normalisation of passives	148/149 (99.3 %)	147/149 (98.7 %)

Compilation of the SCooLE

Annotation

```
I        PP    I      PPIS1 I
'm       VBP   be     VBM    be
not      RB    not    XX     not
<passive constr="beved" oaldadj="no" target="yes">
<replaced meth="manual" orig="allow" type="false">
allowedVVN   allow VVN    allow
</replaced>
</passive>
to       TO    to     TO     to
buy      VV    buy    VVI    buy
it       PP    it     PPH1   it
```

TreeTagger   CLAWS

Evidence from the Corpora

Procedure

Case 1:

TreeTagger and CLAWS agree on classification

of participle:

- *be{0}Ved (be Ved* with no intervening element)
- *be{1}Ved (be Ved* with one intervening element)

⇨ **Automatic analysis of results, no manual check**

Evidence from the Corpora

Procedure

Case 2:
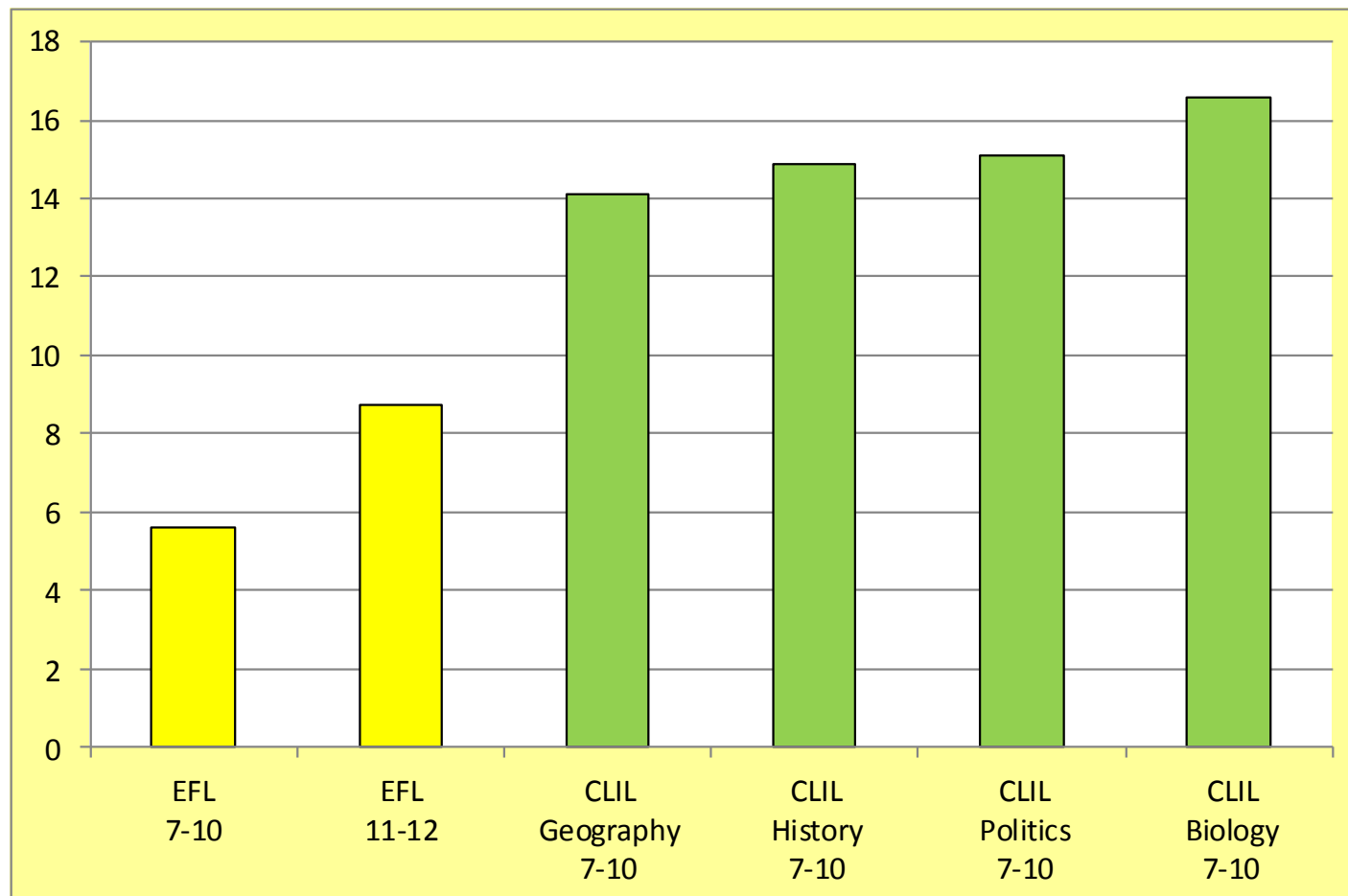
TreeTagger and CLAWS do not agree on classification of participle:

- *be{0}Ved (be Ved* with no intervening element)
- *be{1}Ved (be Ved* with one intervening element)

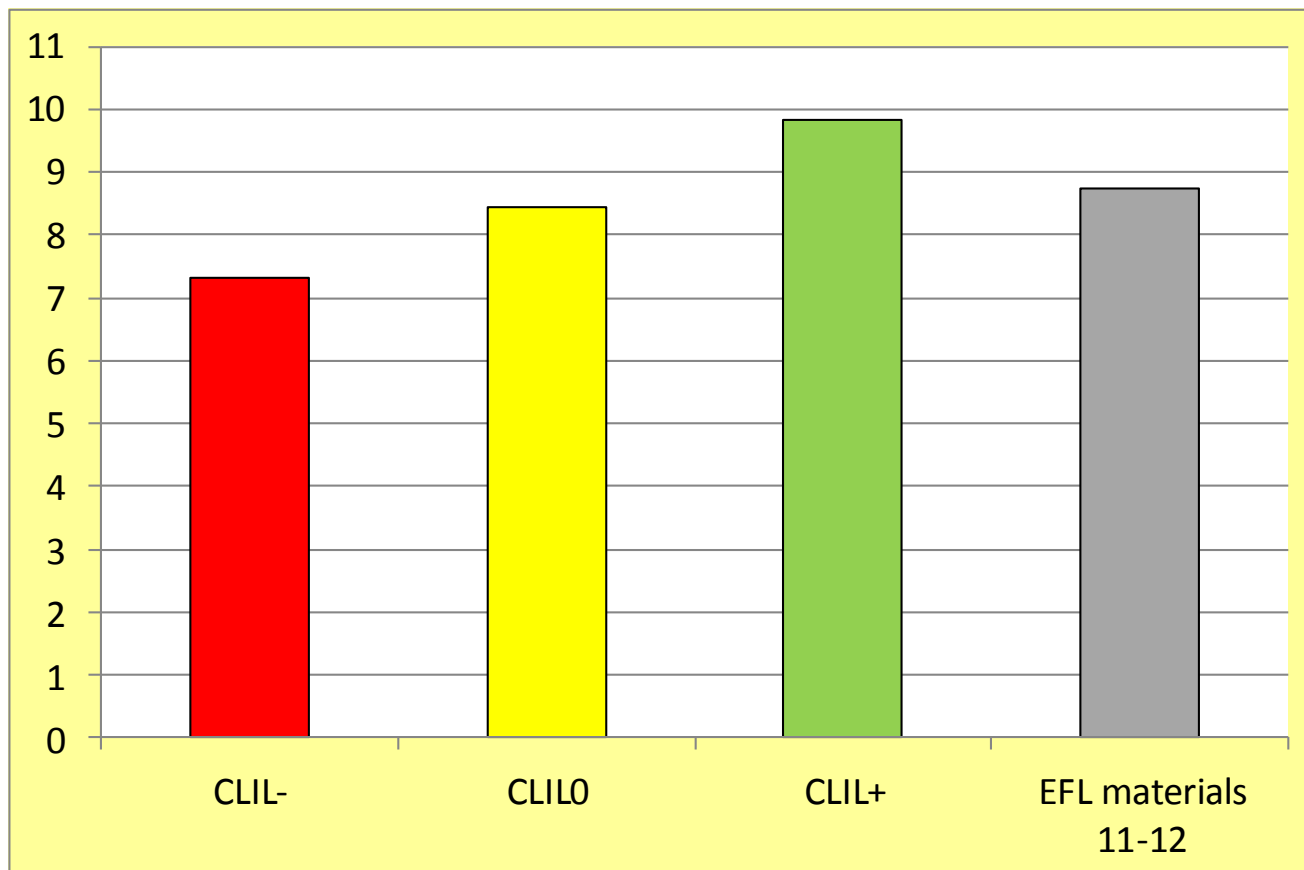⇨ **Automatic analysis of results + manual check**

Evidence from the TeaMC

Frequency: *be{0}Ved* and *be{1}Ved* per 1,000 words - Results

⇨ Relative number of passives is considerably higher in CLIL materials
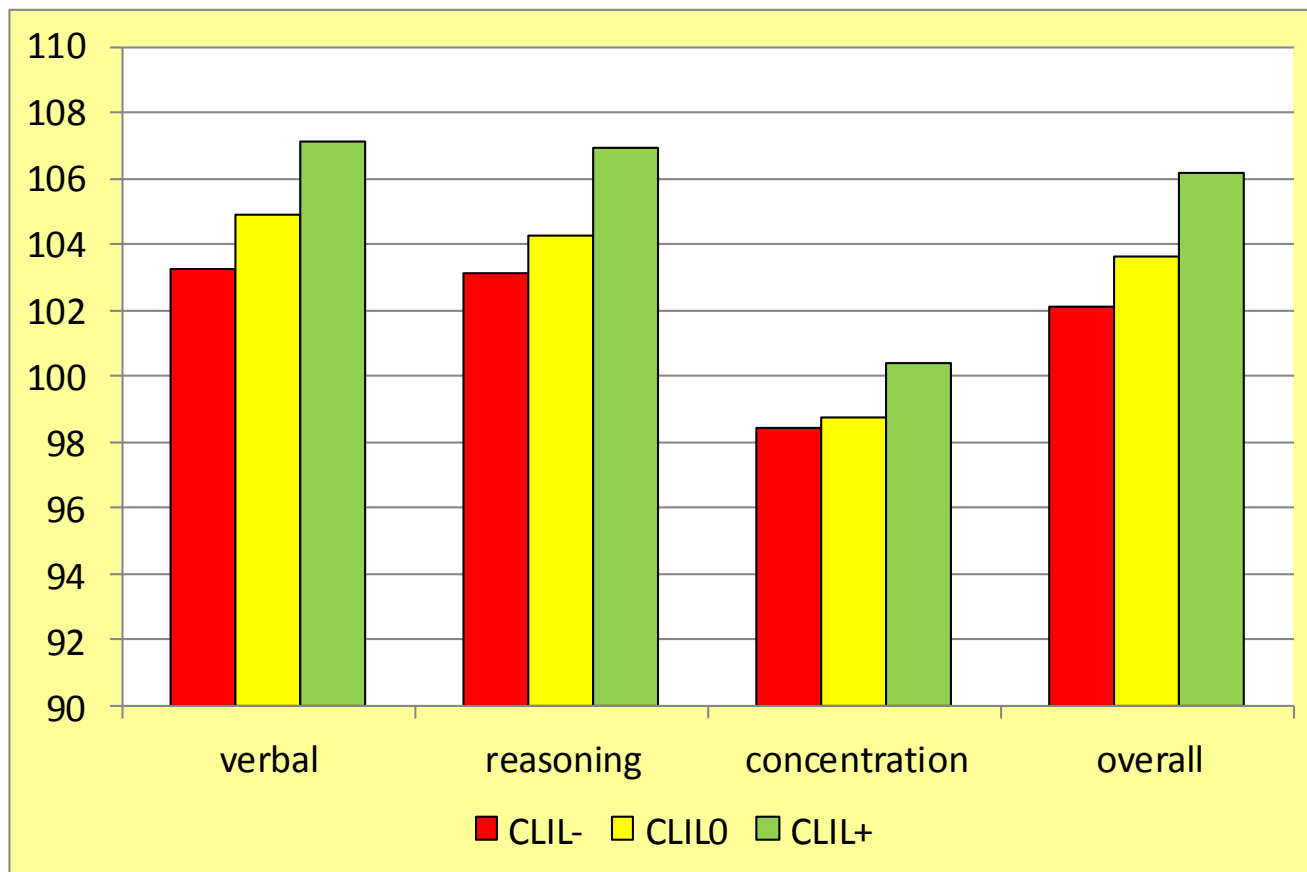
Evidence from the SCooLE

Frequency: *be{0}Ved* and *be{1}Ved* per 1,000 words - Results

⇨ Participants in CLIL produce a higher relative number of passives than are displayed in EFL materials for any level!

Evidence from the SCooLE

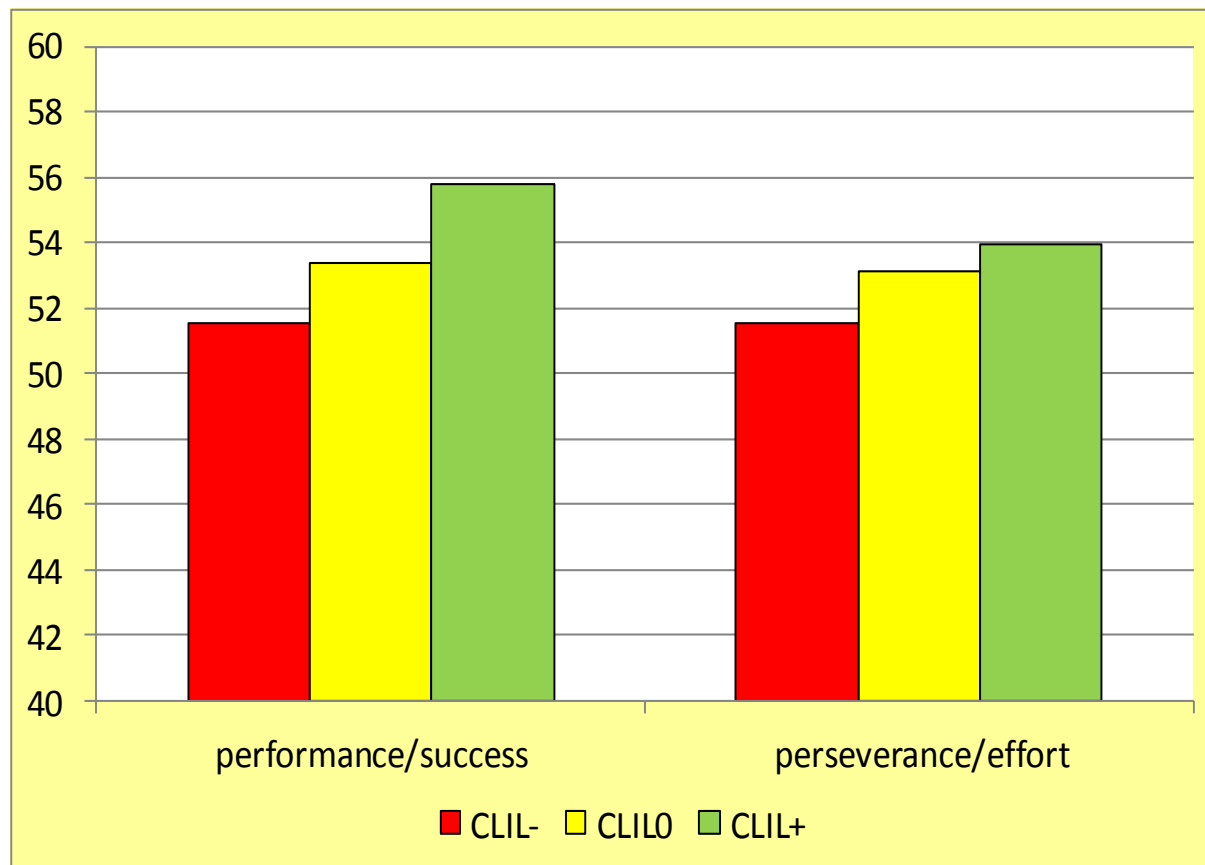Possible influence of learner variables – Intelligence (standard score)

⇨ CLIL programmes have a tendency to sort students into groups of (non-)participants according to their cognitive skills

Evidence from the SCooLE

Possible influence of learner variables – Motivation (T-score)

⇨ CLIL programmes have a tendency to sort students into groups of (non-)participants according to their motivation

Evidence from the SCooLE

Possible influence of learner variables – Consequences

Bruton 2011:

"Is CLIL so beneficial, or just selective?"

⇨ CLIL is indeed **selective**

⇨ Differences in the use of the English passive may be due to **factors other than educational setting**

⇨ Sound statistical analysis of data is needed to determine **the extent to which CLIL is beneficial** with respect to the English passive

Evidence from the SCooLE

Passive ratio

Granger 2013:

"Some verbs display strong passive attraction, while others are characterized by passive repulsion."

Verbs with high passive ratio	
lemma	passive ratio
oblige	68.2 %
deem	60.0 %
entitle	55.3 %
expect	53.5 %

Verbs with low passive ratio	
lemma	passive ratio
learn	4.2 %
receive	3.3 %
attend	2.7 %
want	0.8 %

Evidence from the SCooLE

Passive ratio *(be{0}Ved and be{1}Ved)*

Lemma	SCooLE			TeaMC
	CLIL-	CLIL0	CLIL+	EFL 11-12
raise	40.0 %	36.3 %	31.7 %	18.8 %
allow	92.6 %	84.0 %	77.8 %	20.0 %
(re-)introduce	56.3 %	61.2 %	55.7 %	26.0 %
give	10.5 %	22.3 %	17.9 %	7.9 %
kill	11.4 %	17.7 %	28.7 %	22.2 %
discuss	58.8 %	18.1 %	24.2 %	7.7 %
make	5.3 %	6.0 %	8.3 %	7.0 %
treat	66.7 %	52.9 %	59.5 %	32.4 %
create	44.4 %	40.0 %	36.7 %	12.8 %
see	4.9 %	5.1 %	6.8 %	7.4 %

Most frequent passive forms (SCooLE: all subcorpora)

Evidence from the SCooLE

Passive ratio *(be{0}Ved and be{1}Ved)*

Lemma	SCooLE			TeaMC
	CLIL-	**CLIL0**	**CLIL+**	**EFL 11-12**
raise	40.0 %	36.3 %	31.7 %	18.8 %
allow	92.6 %	84.0 %	77.8 %	20.0 %
(re-)introduce	56.3 %	61.2 %	55.7 %	26.0 %
give	10.5 %	22.3 %	17.9 %	7.9 %
kill	11.4 %	17.7 %	28.7 %	22.2 %
discuss	58.8 %	18.1 %	24.2 %	7.7 %
make	5.3 %	6.0 %	8.3 %	7.0 %
treat	66.7 %	52.9 %	59.5 %	32.4 %
create	44.4 %	40.0 %	36.7 %	12.8 %
see	4.9 %	5.1 %	6.8 %	7.4 %

⇨ Overuse possibly triggered by prompts

Evidence from the SCooLE

Passive ratio *(be{0}Ved and be{1}Ved)*

Lemma	SCooLE			TeaMC
	CLIL-	**CLIL0**	**CLIL+**	**EFL 11-12**
raise	40.0 %	36.3 %	31.7 %	18.8 %
allow	92.6 %	84.0 %	77.8 %	20.0 %
(re-)introduce	56.3 %	61.2 %	55.7 %	26.0 %
give	10.5 %	22.3 %	17.9 %	7.9 %
kill	11.4 %	17.7 %	28.7 %	22.2 %
discuss	58.8 %	18.1 %	24.2 %	7.7 %
make	5.3 %	6.0 %	8.3 %	7.0 %
treat	66.7 %	52.9 %	59.5 %	32.4 %
create	44.4 %	40.0 %	36.7 %	12.8 %
see	4.9 %	5.1 %	6.8 %	7.4 %

⇨ Overuse possibly triggered by introduction
as lexical chunk in EFL materials

Evidence from the SCooLE

Passive ratio *(be{0}Ved and be{1}Ved)*

Lemma	SCooLE			TeaMC
	CLIL-	**CLIL0**	**CLIL+**	**EFL 11-12**
raise	40.0 %	36.3 %	31.7 %	18.8 %
allow	92.6 %	84.0 %	77.8 %	20.0 %
(re-)introduce	56.3 %	61.2 %	55.7 %	26.0 %
give	10.5 %	22.3 %	17.9 %	7.9 %
kill	11.4 %	17.7 %	28.7 %	22.2 %
discuss	58.8 %	18.1 %	24.2 %	7.7 %
make	5.3 %	6.0 %	8.3 %	7.0 %
treat	66.7 %	52.9 %	59.5 %	32.4 %
create	44.4 %	40.0 %	36.7 %	12.8 %
see	4.9 %	5.1 %	6.8 %	7.4 %

⇨ Overuse possibly triggered by transfer from German

Evidence from the SCooLE

Passive ratio *(be{0}Ved and be{1}Ved)*

Lemma	SCooLE			TeaMC
	CLIL-	CLIL0	CLIL+	EFL 11-12
raise	40.0 %	36.3 %	31.7 %	18.8 %
allow	92.6 %	84.0 %	77.8 %	20.0 %
(re-)introduce	56.3 %	61.2 %	55.7 %	26.0 %
give	10.5 %	22.3 %	17.9 %	7.9 %
kill	11.4 %	17.7 %	28.7 %	22.2 %
discuss	58.8 %	18.1 %	24.2 %	7.7 %
make	5.3 %	6.0 %	8.3 %	7.0 %
treat	66.7 %	52.9 %	59.5 %	32.4 %
create	44.4 %	40.0 %	36.7 %	12.8 %
see	4.9 %	5.1 %	6.8 %	7.4 %

⇨ Underuse possibly triggered by avoidance
   in less proficient learners

Conclusion

1. Participants in CLIL programmes use a higher relative number of passives.

This may be due to

- the type of input they are faced with in CLIL;
- learner variables.

Future work:

Statistical analysis of available data to determine the exact influence of educational setting.

Conclusion

2. All groups of learners overuse the passive with respect to the passive ratio of some verbs while underusing it with respect to others.

This may, amongst others, be due to
- the treatment of passives by EFL materials;
- transfer from L1.

Future work:
- Analysis of the representation of passives in EFL materials;
- analysis of corresponding verbs in L1.

References

Baron, Alistair & Rayson, Paul (2009). Automatic standardisation of texts containing spelling variation. How much training data do you need? In: Michaela Mahlberg, Victorina González-Díaz & Catherine Smith [Eds.]. Proceedings of the Corpus Linguistics Conference, CL 2009, Liverpool, UK, 2009.

Bruton, Anthony (2011). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. In: System, 39, 2011. 523-531.

Evert, Stefan & Hardie, Andrew (2011). Twenty-first century corpus workbench: Updating a query architecture for the new millennium. In: Proceedings of the Corpus Linguistics 2011 Conference, Birmingham, UK.

Garside, Roger & Smith, Nicolas (1997). A hybrid grammatical tagger: CLAWS4. In: Roger Garside, Geoffrey Leech & Anthony McEnery [Eds.]. Corpus Annotation: Linguistic Information from Computer Text Corpora. London: Longman. 102-121.

Granger, Sylviane (2013). The passive in learner English. Corpus insights and implications for pedagogical grammar. In: Ishikawa, Shin Ichiro [Ed.]. Learner Corpus Studies in Asia and the World. Vol. 1. Papers from LCSAW2013 (5-15). Kobe: School of Languages and Communication, Kobe University.

Holtz, Mônica (2011). Lexico-grammatical properties of abstracts and research articles. A corpus-based study of scientific discourse from multiple disciplines. Darmstadt: Technische Universität, PhD Thesis. Tuprints.ulb.tu-darmstadt.de/2638/1/PhD-Thesis-Monica-Holtz.pdf

Horn, Wolfgang (2003). PSB-R 6-13. Prüfsystem für Schul- und Bildungsberatung für 6. bis 13. Klassen – revidierte Fassung. Göttingen: Hogrefe.

Petermann, Franz & Winkel, Sandra (2007). FLM 7-13. Fragebogen zur Leistungsmotivation für Schüler der 7. bis 13. Klasse. Frankfurt/Main: Harcourt.

Schmid, Helmut (1994): Probabilistic Part-of-Speech Tagging Using Decision Trees. In: Proceedings of International Conference on New Methods in Language Processing, Manchester, UK.

Svartvik, Jan (1966). On Voice in the English Verb. The Hague/Paris: Mouton.

Wanner, Anja (2009). Deconstructing the English passive. Berlin: Mouton de Gruyter.

Möller, Verena (2013). How Do Educational Settings at the Secondary Level Impact on Learners' Use of the English Passive? – Evidence from the Secondary-Level Corpus of Learner English (SCooLE). Paper presented at *Learner Corpus Research, LCR 2013,* Bergen/Norway, 27-29 September 2013.