



# Corpus-driven analysis of cotextual patterns in learner Finnish

Jarmo H. Jantunen

Learner Corpus Colloquium  
Centre for English Corpus Linguistics  
University of Louvain, Belgium  
10 September, 2007



# Project

- **Corpus study on language-specific and universal features in learner language (2007–)**
- Joint project: University of Oulu (Finland), Univ. of Umeå (Sweden), Univ. of Tallinn (Estonia), State Univ. of Petrozavodsk (Russia)
- Corpus: International Corpus of Learner Finnish (ICLFI)
  - 26 foreign universities
  - Learner Finnish from 15 different mother tongues
  - August 2007: 25 000 tokens (Swedish, German)



# Key word analysis NNF vs. NF (WordSmith Tools, Mike Scott)

Step 1: Word lists of NNF and NF

Step 2: Statistical comparison of the word lists of NNF and NF

N	TOKEN		f / NNF Corpus	f / NF Corpus	keyness
1	TIMO	[Timo-NOM]	130	114	976,6
2	ANA	[Ana-NOM]	64	1	635,6
3	KÖLN	[Köln-NOM]	60	0	605,5
4	KÖLNIN	[Köln-GEN]	50	6	466,5
5	TOIVONEN	[Toivonen-NOM]	47	15	405,9
6	KALEVALA	[Kalevala-NOM]	44	8	399,5
7	TÄNÄÄN	[’today]	83	495	368,5
8	KELLO	[watch, o’clock]	79	435	361,8
9	SYÖN	[eat-1sg]	43	29	337,2
10	OSTAISIN	[buy-1sg-cond]	33	6	299,6
.					
.					
.					



# KELLO 'watch, o'clock' in NNF and NF: cotext analysis

	NNF	NF
Atypically frequent in NNF:	33 / 10.000	2 / 10.000
Atypical collocations in NNF:	<p><u>noin</u> 'about' kello 11</p> <p><u>herätä</u> 'to weak up' kello 9</p> <p>kello <u>olla</u> 'to be' 8</p> <p>kello <u>kahdeksan</u> 'eight' 7</p> <p>kello <u>kuusi</u> 'six' 7</p> <p>kello <u>puoli</u> 'half' 6</p> <p>kello <u>kymmenen</u> 'ten' 5</p>	<p>kello <u>olla</u> 'to be' 144</p> <p>kello <u>ja</u> 'and' 86</p> <p><u>vilkaista</u> 'to glance at' kello 58</p> <p><u>katsoa</u> 'to check' kello 36</p> <p><u>kun</u> 'when' kello 19</p> <p>kello <u>viisi</u> 'five' 19</p> <p>kello <u>yhdeksän</u> 'nine' 19</p>
Semantic preference KELLO <'time'> Atypically high proportion in NNF:	85 % (of all occurrences) (e.g. <i>Heräsin kello yhdeksältä.</i> 'I woke up at nine o'clock.')	29 % (of all occurrences)
Semantic preference KELLO <'to look at'> Atypically low proportion in NNF:	0 % (of all occurrences)	14 % (of all occurrences) ( <i>Vilkaisin kelloa, se oli tasan yhdeksän.</i> 'I glanced at the watch, it was nine sharp.'



# Tentative hypotheses for further research

Hypotheses on atypicalities in learner language:

- atypical frequencies of lexical items (quant.)
- atypical frequencies of cotextual patterns (quant.)
- atypical cotextual patterns - *also other than collocations, i.e. colligations and semantic preferences* (qualit.)
  
- atypical cotextual patterning is a widespread tendency in learner language – perhaps even *a learner language universal*.

