

I haven't drank in weeks: The use of past tense forms as past participles in English corpora

Kristina Geeraert
University of Alberta
AACL – Oct. 9, 2009

Wolfram (2003)

- “past for perfect as in *they had went there*, occurs in socially subordinate varieties of English wherever they are found throughout the world.” (p. 146)

Cf. also Cheshire 1982, 1994, Eisikovits 1987, Lass 1994, Wright 1981, etc.

Oxford English Dictionary

- <took>
 - For the past participle *taken*, the past tense *took* has been common since the 16th c. in vulgar speech and in dialects
- <drank>
 - On the other hand, from 17th to 19th c. *drank* was intruded from the past tense into the past participle, prob. to avoid the inebriate associations of *drunk*.
- <went>
 - The use of *went* as a past participle is sometimes heard in illiterate speech
- <beat>
 - The past participle *beat*, still occasional for *beaten* in all senses, but chiefly used in sense 10, ...

Structure of Talk

- Previous research
- Use BNC and COCA to investigate:
 - Syntactic context
 - Genre
 - Varieties of English
- Web as corpus
- Conclusion

Strong Verb Paradigm

(adapted from Anderwald 2009)

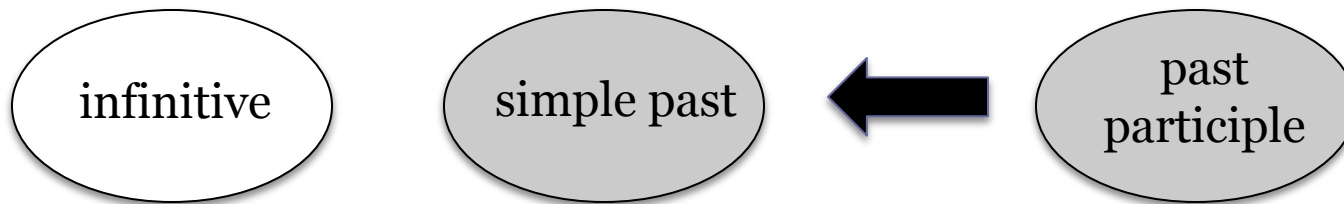
(a)	infinitive 'cut'	simple past 'cut'	past participle 'cut'	14.4%
(b)	infinitive 'meet'	simple past 'met'	past participle 'met'	48.5%
(c)	infinitive 'beat'	simple past 'beat'	past participle 'beaten'	0.6%
(d)	infinitive 'come'	simple past 'came'	past participle 'come'	1.2%
(e)	infinitive 'drink'	simple past 'drank'	past participle 'drunk'	35.3%
				<hr/> 100%

Strong Verb Paradigm

(adapted from Anderwald 2009)

(a)	infinitive 'cut'	simple past 'cut'	past participle 'cut'	14.4%
(b)	infinitive 'meet'	simple past 'met'	past participle 'met'	48.5%
(c)	infinitive 'beat'	simple past 'beat'	past participle 'beaten'	0.6%
(d)	infinitive 'come'	simple past 'came'	past participle 'come'	1.2%
(e)	infinitive 'drink'	simple past 'drank'	past participle 'drunk'	35.3%
				<hr/> 100%

Bybee Verbs (Bybee 1982, 1995, Anderwald 2007, 2009)



C (C) (C) V {velar/nasal**}**

dig ~ dug

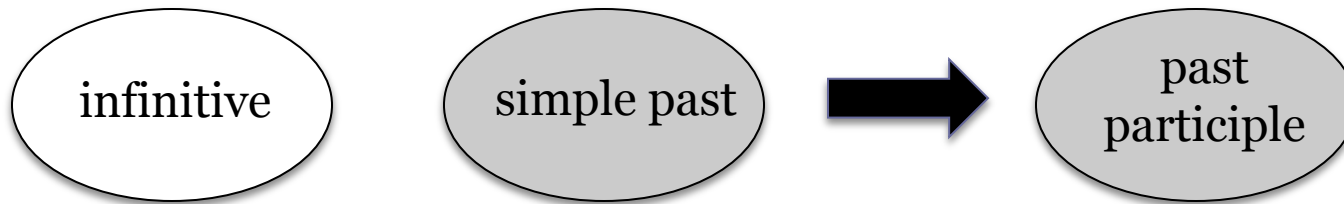
swim ~ swum

drink ~ drunk

cling ~ clung

sting ~ stung

“Geeraert” Verbs



NO common phonetic structure

beat ~ beat (beaten)
take ~ took (taken)
show ~ showed (shown)
go ~ went (gone)

bite ~ bit (bitten)
fall ~ fell (fallen)
ring ~ rang (rung)
come ~ came (come)

Strong Verb Paradigm

(adapted from Anderwald 2009)

(a)	infinitive 'cut'	simple past 'cut'	past participle 'cut'	14.4%
(b)	infinitive 'meet'	simple past 'met'	past participle 'met'	48.5%
(c)	infinitive 'beat'	simple past 'beat'	past participle 'beaten'	0.6%
(d)	infinitive 'come'	simple past 'came'	past participle 'come'	1.2%
(e)	infinitive 'drink'	simple past 'drank'	past participle 'drunk'	35.3%
				<hr/> 100%

Strong Verb Paradigm

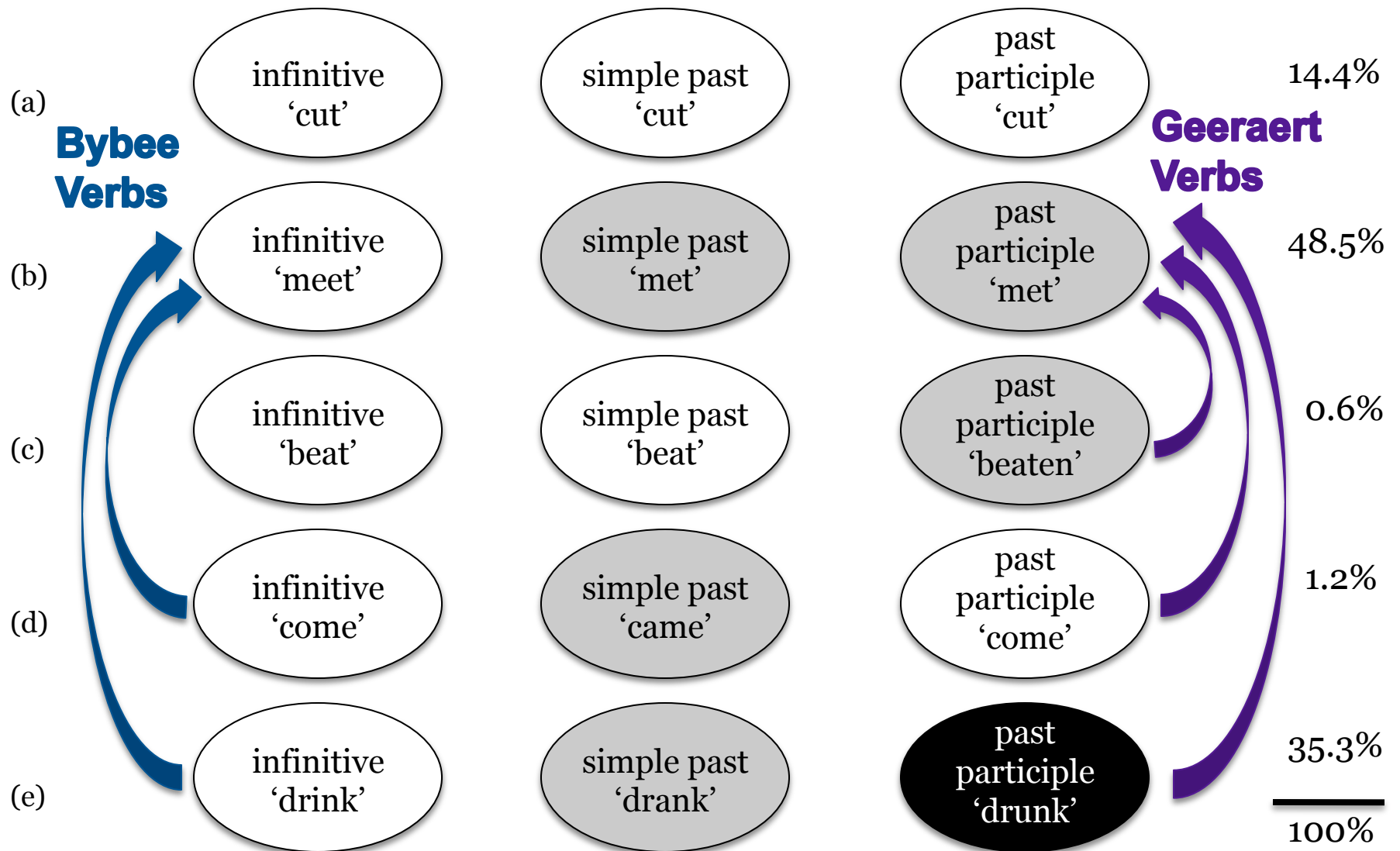
(adapted from Anderwald 2009)

(a)	infinitive 'cut'	simple past 'cut'	past participle 'cut'	14.4%
(b)	infinitive 'meet'	simple past 'met'	past participle 'met'	48.5%
(c)	infinitive 'beat'	simple past 'beat'	past participle 'beaten'	0.6%
(d)	infinitive 'come'	simple past 'came'	past participle 'come'	1.2%
(e)	infinitive 'drink'	simple past 'drank'	past participle 'drunk'	35.3%
				<hr/> 100%

Bybee Verbs

Strong Verb Paradigm

(adapted from Anderwald 2009)



Structure of Talk

- Previous research
- Use BNC and COCA to investigate:
 - Syntactic context
 - Genre
 - Varieties of English
- Web as corpus
- Conclusion

BNC (British National Corpus, 100 million words)

	<HAVE>	<HAS>	<HAD>	<VE>	<HAVEN'T>	<HASN'T>	<HADN'T>
ATE	0	0	0	5	3	0	0
BEAT	3	0	5	3	0	0	0
BIT	4	0	3	2	1	0	0
CAME	8	0	5	5	0	0	0
DRANK	0	0	1	2	0	0	0
FELL	5	1	4	4	0	0	0
GOT	3257	1425	1367	26505	2618	509	364
RAN	5	0	3	0	0	0	1
RANG	4	1	2	5	1	0	3
RODE	2	0	0	1	0	0	1
SANG	0	0	0	0	0	0	0
SANK	2	0	0	0	0	0	0
SHOWED	13	6	30	3	3	1	2
SPOKE	1	1	0	8	3	2	0
TOOK	31	4	2	20	0	0	0
WENT	5	1	3	5	0	0	0
WROTE	9	0	3	3	0	1	0

BNC (British National Corpus, 100 million words)

	<HAVE>	<HAS>	<HAD>	<VE>	<HAVEN'T>	<HASN'T>	<HADN'T>
ATE	0	0	0	5	3	0	0
BEAT	3	0	5	3	0	0	0
BIT	4	0	3	2	1	0	0
CAME	8	0	5	5	0	0	0
DRANK	0	0	1	2	0	0	0
FELL	5	1	4	4	0	0	0
GOT	3257	1425	1367	26505	2618	509	364
RAN	5	0	3	0	0	0	1
RANG	4	1	2	5	1	0	3
RODE	2	0	0	1	0	0	1
SANG	0	0	0	0	0	0	0
SANK	2	0	0	0	0	0	0
SHOWED	13	6	30	3	3	1	2
SPOKE	1	1	0	8	3	2	0
TOOK	31	4	2	20	0	0	0
WENT	5	1	3	5	0	0	0
WROTE	9	0	3	3	0	1	0

COCA (Corpus of Contemporary American English, 400+ million words)

	<HAVE>	<HAS>	<HAD>	<'VE>	<HAVEN'T>	<HASN'T>	<HADN'T>
ATE	0	0	3	2	0	1	0
BEAT	46	20	67	20	3	1	2
BIT	12	2	16	3	2	0	0
CAME	48	6	44	7	2	0	1
DRANK	10	1	6	5	6	0	1
FELL	6	1	16	2	0	0	2
GOT	3678	2986	652	63234	766	245	103
RAN	11	4	18	7	0	0	0
RANG	0	0	2	0	1	0	0
RODE	2	0	2	5	0	2	0
SANG	6	0	5	3	0	0	0
SANK	2	1	0	0	0	0	0
SHOWED	92	55	124	34	3	6	12
SPOKE	7	5	11	12	2	0	2
TOOK	38	2	22	15	0	0	1
WENT	81	10	66	26	3	1	1
WROTE	12	2	10	3	1	0	0

COCA (Corpus of Contemporary American English, 400+ million words)

	<HAVE>	<HAS>	<HAD>	<'VE>	<HAVEN'T>	<HASN'T>	<HADN'T>
ATE	0	0	3	2	0	1	0
BEAT	46	20	67	20	3	1	2
BIT	12	2	16	3	2	0	0
CAME	48	6	44	7	2	0	1
DRANK	10	1	6	5	6	0	1
FELL	6	1	16	2	0	0	2
GOT	3678	2986	652	63234	766	245	103
RAN	11	4	18	7	0	0	0
RANG	0	0	2	0	1	0	0
RODE	2	0	2	5	0	2	0
SANG	6	0	5	3	0	0	0
SANK	2	1	0	0	0	0	0
SHOWED	92	55	124	34	3	6	12
SPOKE	7	5	11	12	2	0	2
TOOK	38	2	22	15	0	0	1
WENT	81	10	66	26	3	1	1
WROTE	12	2	10	3	1	0	0

<Modal + *have* + Past Tense/Participle>

- *could, would, should, ought to, must, might, etc.*
 - *It **must have took** about three years (BNC)*
 - *My father **could have went** to jail (COCA)*
 - *It **should not have came** to this (COCA)*
 - ***Could** one person **have ran** a massive scheme? (COCA)*
- counted all modals occurring within 5 positions to the left of *have*

<modal *have went*> in COCA

<u>modal <i>have went</i></u>	= <u>67</u>	= 82.7%
all instances of <i>have went</i>	81	

<u>modal <i>have gone</i></u>	= <u>3623</u>	= 40.9%
all instances of <i>have gone</i>	8860	

Cf. Eisikovits 1987

Percent difference of <modal *have went*> in COCA

	<have went>	<have gone>	Total
Modal	67 (82.7%)	3623 (40.9%)	3690
No Modal	14 (17.3%)	5237 (59.1%)	5251
Total	81	8860	

%d = 41.8%

Percent differences in use of <Modal + *have* + Past Tense> in COCA

- <have saw>
 %*d* = 60.7%
- <have drove>
 %*d* = 60.3%
- <have fell>
 %*d* = 59.8%
- <have swam>
 %*d* = 59.1%
- <have ran>
 %*d* = 50%
- <have gave>
 %*d* = 46.1%
- <have came>
 %*d* = 43.2%
- <have went>
 %*d* = 41.8%
- <have wrote>
 %*d* = 40.4%
- <have took>
 %*d* = 39.7%
- <have showed>
 %*d* = 32.1%
- <have broke>
 %*d* = 31.3%
- <have sang>
 %*d* = 31%
- <have beat>
 %*d* = 30.5%
- <have began>
 %*d* = 19.2%
- <have drank>
 %*d* = 6.7%

Genres

	BNC		COCA	
SPOKEN	154	56.9%	648	53.2%
FICTION	42	15.5%	300	24.6%
NEWSPAPER	15	5.5%	134	11.0%
ACADEMIC	15	5.5%	50	4.1%
MAGAZINE	–	–	86	7.1%
MISC	45	16.6%	–	–
	271	100%	1218	100%

Genres

	BNC		COCA	
	SPOKEN	154	56.9%	648
FICTION	42	15.5%	300	24.6%
NEWSPAPER	15	5.5%	134	11.0%
ACADEMIC	15	5.5%	50	4.1%
MAGAZINE	–	–	86	7.1%
MISC	45	16.6%	–	–
	271	100%	1218	100%

BNC vs. COCA

<*have + did*>

- has past participle usage only in COCA

*I would **have did** anything for her*

*You shouldn't **have did** that*

- <HAVE DID> 53 (total of 61 instances)
- <HAS DID> 3 (total of 3 instances)
- <HAD DID> 25 (total of 49 instances)
- <'VE DID> 8 (total of 9 instances)

Structure of Talk

- Previous research
- Use BNC and COCA to investigate:
 - Syntactic context
 - Genre
 - Varieties of English
- **Web as corpus**
- Conclusion

Chat Corpus

- Corpus Eye -
<http://corp.hum.sdu.dk/cqp.en.html>
- 23.5 million words
- Compiled from internet chat rooms
- Reproducible
- % = non-standard verb
sum of standard & non-standard

Chat Corpus

NON-STANDARD	PERCENTAGE	NON-STANDARD	PERCENTAGE
HAVE WENT - 22	8.7%	HAVE TOOK - 5	2.5%
HAS WENT - 10	3.9%	HAS TOOK - 3	1.9%
HAD WENT - 15	11.7%	HAD TOOK - 2	1.2%
HAVE ATE - 6	24%	HAVE CAME - 25	11.5%
HAVE N'T ATE - 7	16.3%	HAS CAME - 3	1.8%
HAS N'T ATE - 2	4%	HAD CAME - 95	43.8%
HAD N'T ATE - 2	15.4%	HAS N'T CAME - 3	11.5%
HAVE RAN - 13	37.1%	HAVE BEAT - 6	40%
HAS RAN - 6	20.7%	HAS BEAT - 4	44.4%
HAD RAN - 8	28.6%	HAD BEAT - 2	50%
HAVE SHOWED - 11	25.6%	HAVE WROTE - 3	25%
HAS N'T SHOWED - 2	22.2%	HAD WROTE - 2	11.8%

Google Results

- **low** percentages of some non-standard verbs in BNC & COCA also show relatively **low** percentages in Google
- % = non-standard verb
sum of standard & non-standard

NON-STANDARD	.com		.ca		.uk	
HAVE CAME	813,000	1.9%	24,100	1.2%	32,500	1.6%
HAVE SHOWED	312,000	0.8%	15,200	2.1%	20,500	2.1%
HAVE TOOK	226,000	0.7%	9,130	0.6%	14,300	0.7%
HAVE WENT	830,000	2.9%	33,900	2.7%	16,700	1.0%
HAVE WROTE	226,000	1.4%	11,400	1.6%	9,450	0.9%

Google Results

- **high** percentages of some non-standard verbs in BNC & COCA also show relatively **high** percentages in Google

NON-STANDARD	.com		.ca		.uk	
HAVE BEAT	683,000	46.1%	26,700	44.7%	31,700	23.4%
HAVE BIT	74,400	39.9%	823	35.3%	3,530	23.6%

Google Results

- **low** percentages of some non-standard verbs in BNC & COCA may actually show relatively **high** percentages in Google

NON-STANDARD	.com		.ca		.uk	
HAVE DRANK	98,000	29.1%	3,900	23.1%	6,290	21.7%
HAVE RANG	23,700	27.4%	696	49.1%	3,880	23.5%
HAVE RODE	48,600	13.9%	830	14.0%	5,670	21.8%
HAVE SANK	69,700	21.5%	547	12.4%	4,280	10.4%
HAVE SWAM	22,500	46.3%	290	41.0%	937	13.5%

Derived Verbs

	.com		.ca		.uk	
<have befell>	3,320	5.0%	40	2.9%	117	2.3%
<have forgot>	391,000	9.0%	709	1.2%	7,020	1.6%
<have mistook>	32,700	17.5%	212	29.6%	1,270	8.0%
<have overcame>	4,890	0.8%	189	1.2%	543	1.1%
<have reran>	1,370	6.8%	0	0.0%	3	2.9%
<have rewrote>	13,200	13.7%	91	6.4%	291	3.5%
<have undertook>	7,530	1.4%	323	0.7%	986	0.3%
<have underwent>	28,400	2.2%	385	0.7%	859	0.4%

- % = non-standard verb
sum of standard & non-standard
- Non-standard forms less common with derived verbs

Derived Verbs

	.com		.ca		.uk	
<have befall>	3,320	5.0%	40	2.9%	117	2.3%
<have forgot>	391,000	9.0%	709	1.2%	7,020	1.6%
<have mistook>	32,700	17.5%	212	29.6%	1,270	8.0%
<have overcame>	4,890	0.8%	189	1.2%	543	1.1%
<have reran>	1,370	6.8%	0	0.0%	3	2.9%
<have rewrote>	13,200	13.7%	91	6.4%	291	3.5%
<have undertook>	7,530	1.4%	323	0.7%	986	0.3%
<have underwent>	28,400	2.2%	385	0.7%	859	0.4%

- % = non-standard verb
sum of standard & non-standard
- Non-standard forms less common with derived verbs

Conclusions

- Not a recent colloquial development
- “Stable low-frequency” phenomenon
 - (cf. Anderwald 2009)
- More frequent usage of modal auxiliaries with non-standard forms
- 50% usage is spoken
- Derived verbs less likely to be regularized
- “Web as corpus” – different behaviour, not just larger numbers
 - (cf. Hundt et al. 2007)



THANK YOU!

Acknowledgements
John Newman

References

- Anderwald, L. (2009). *The morphology of English dialects: Verb formation in non-standard English*. Cambridge: Cambridge University Press.
- Anderwald, L. (2007). 'He rung the bell' and 'she drunk ale' – non-standard past tense forms in traditional British dialects and on the internet. In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), *Corpus linguistics and the web* (pp. 271-286). Amsterdam, NL: Rodopi.
- Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* 10, 425-455.
- Bybee, J. L. & D. I. Slobin. (1982). Why small children cannot change language on their own: suggestions from the English past tense. In A. Ahlqvist (Ed.), *Papers from the 5th international conference on historical linguistics* (pp. 29-37). Amsterdam, NL: John Benjamins.
- Chesire, J. (1994). Standardization and the English irregular verbs. In D. Stein & I. Tieken-Boon van Ostade (Eds.), *Towards a standard English: 1600-1800* (pp. 115-133). Berlin: Mouton de Gruyter.

References

- Chesire, J. (1982). *Variation in an English dialect: A sociolinguistic study*. Cambridge: Cambridge University Press.
- Eisikovits, E. (1987). Variations in the lexical verb in inner-Sydney English. *Australian Journal of Linguistics* 7, 1-24.
- Hundt, M., Nesselhauf, N. & Biewer, C. (2007). *Corpus linguistics and the web*. Amsterdam, NL: Rodopi.
- Lass, R. (1994). Proliferation and option-cutting: the strong verb in the fifteenth to eighteenth centuries. In D. Stein & I. Tieken-Boon van Ostade (Eds.), *Towards a standard English: 1600-1800* (pp. 81-113). Berlin: Mouton de Gruyter.
- *Oxford English Dictionary* (OED). (2009). Retrieved on May 16, 2009 from <http://dictionary.oed.com.login.ezproxy.library.ualberta.ca>.
- Wolfram, W. (2003). Enclave dialect communities in the south. In S. Nagle & S. Sanders (Eds.), *English in the southern United States* (pp. 141-158). Cambridge: Cambridge University Press.
- Wright, P. (1981). *Cockney dialect and slang*. London: B. T. Batsford.