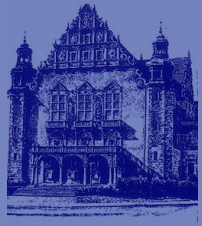# Vagueness tags
in Polglish & Polish

Joanna JENDRYCZKA-WIERSZYCKA

Adam Mickiewicz University, Poznań, Poland

Contents

- Objectives of the study

- Methodological framework

- Data preprocessing

- Vagueness tags

- 'sort of' in English

- 'sort of' – Polish equivalents

- Conclusions

- References

Objectives of the study

- Learn about the features of advanced Polglish spoken language

- Present the findings in comparison with Frenglish corpus data

- Focus: vagueness tags, e.g. *you know, I mean, sort of like*

- Investigate possible L2 learner strategies of Polglish speakers

- Find out about L1 VTs

Methodological framework

- LINDSEI Polish, French and native English subcorpora used (PLINDSEI, FLINDSEI and LOCNEC); Polish AOK corpus for CA purpose

- Biber et al.'s (1999) corpus-driven 'recurrent word combination' method of investigating vocabulary use

- the search pertains to two-, three-, four-, five- and six-word sequences occurring at least 12, 6, 4, 3 and 3 times respectively in all four corpora (Altenberg 1990).

- CIA = Contrastive Interlanguage Analysis (Granger 1996) used for comparing interlanguage data

L2 data preprocessing (1)

- B-turns only

<B> don't know if [uh it <\B>

<A>   [okay <\A>

<B> it's the real <begin laughter> <name> <end laughter>
    and say talking about himself [or just <\B>

<A>   [uhu <\A>

- Corpora used in their normalized word number

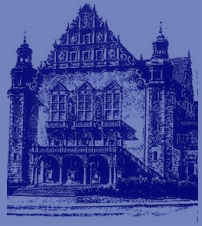| corpus	PL	NS	FR
tokens	90871	90945	90851

L2 (&L1) data preprocessing (2)

- *Cluster* tool of the *WordSmith Tools* software (Scott 1998) used for extracting bundles

- *KeyWords* tool of the *WordSmith* software used for counting overuse and underuse of separate bundles

- Most frequent Polglish clusters were grouped according to structural Biber et al.'s (1999: 996) and functional Biber et al. (1999) and De Cock (2004) categories

Functional categories

- markers of vagueness
- ideational
- markers of attitudinal stance
- responses
- markers of epistemic stance
- quantifying sequences
- markers of speech/thought reporting
- markers of time/place
- topic-dependent sequences

- obligation
- exemplifiers

Vagueness tags (1)

- 30 most frequent Polglish clusters compared against most frequent English bundles

- De Cock's (1998 and 2004) most underused functional group, vagueness tags, compared against the Polish data

- Lack of vagueness tags as learner feature (Channell 1994) and (Crystal – Davy 1975: 111f)

- vague language is regarded by Crystal – Davy (1975: 111f) as "one of the most important features of the vocabulary of informal conversation."

Vagueness tags (2)

- The most important reasons for lexical vagueness, Crystal – Davy (1975: 112)

(a) memory loss – the speaker forgets a word or it may be 'on the tip of his tongue';

(b) lack of a precise equivalent of a word the speaker wants to use in a language spoken, or the speaker's unawareness of the equivalent;

(c) the conversational topic does not require precision;

(d) the vague lexical item is consciously chosen in order to keep an informal atmosphere of the moment.

'sort of' (1)

- Most underused VTs in Frenglish as ground for comparison with Polglish:
    - ➣ overall underuse of VTs by learners
    - ➣ *sort of* has the greatest negative keyness in Polglish!
- 5 out of 35 most underused three-word sequences contain the sequence *sort of*
- See the WST screenshot:

KeyWords - [3pl_ns.kws (keyness)]
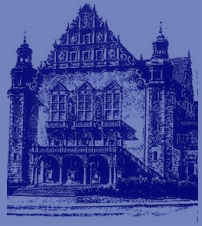
File   Settings   Window   Help

N	WORD	FREQ.	RDB.LS7 %	FREQ.	NS50B.TXT %	KEYNESS	P
30	FIRST OF ALL	15	0,02	0		24,8	0,000001
31	SHE DOESN'T LIKE	0		21	0,02	24,2	0,000001
32	BIT OF A	0		21	0,02	24,2	0,000001
33	I DON'T REALLY	0		21	0,02	24,2	0,000001
34	IT'S A BIT	0		22	0,02	25,3	0,000000
35	I CAN'T REMEMBER	0		22	0,02	25,3	0,000000
36	STUFF LIKE THAT	0		22	0,02	25,3	0,000000
37	WAS A BIT	0		22	0,02	25,3	0,000000
38	THOUGHT IT WAS	0		23	0,02	26,5	0,000000
39	WANTED TO DO	0		23	0,02	26,5	0,000000
40	I MEAN IT	0		23	0,02	26,5	0,000000
41	IN THE MORNING	0		23	0,02	26,5	0,000000
42	TO SORT OF	0		23	0,02	26,5	0,000000
43	YEAH I	0		24	0,02	27,6	0,000000
44	I MEAN IT'S	0		24	0,02	27,6	0,000000
45	END OF THE	0		24	0,02	27,6	0,000000
46	LIKE TO GO	0		24	0,02	27,6	0,000000
47	A LOT MORE	0		24	0,02	27,6	0,000000
48	THAT SORT OF	0		24	0,02	27,6	0,000000
49	WAS REALLY GOOD	0		24	0,02	27,6	0,000000
50	ALL HER FRIENDS	0		25	0,02	28,8	0,000000
51	A SORT OF	0		25	0,02	28,8	0,000000
52	THINGS LIKE THAT	8		56	0,05	29,5	0,000000
53	SO THAT WAS	0		28	0,02	32,2	0,000000
54	BUT I MEAN	0		31	0,03	35,7	0,000000
55	AND THINGS LIKE	0		31	0,03	35,7	0,000000
56	YEAH THAT'S RIGHT	0		32	0,03	36,8	0,000000
57	I'D LIKE TO	6		62	0,05	40,7	0,000000
58	THAT'S RIGHT YEAH	0		37	0,03	42,6	0,000000
59	YEAH YEAH	0		40	0,03	46,0	0,000000
60	SORT OF THING	0		44	0,04	50,6	0,000000
61	I MEAN I	8		83	0,07	54,6	0,000000
62	AT THE MOMENT	0		49	0,04	56,4	0,000000
63	A BIT OF	0		49	0,04	56,4	0,000000
64	SORT OF LIKE	0		50	0,04	57,6	0,000000
65	YEAH YEAH	0		67	0,06	77,1	0,000000

'sort of' - categories

- Polish categories of the sequence 'sort of' *generally* correspond to Aijmer's (2002) categories for native usage:

- 1. *sort of* = "type of" e.g. "a robin is a sort of a bird"

- 2. *sort of* = "approximately", "roughly" or "rather" (Aijmer's core meaning of *sort of* and Crystal – Davy's 1975 "lack of a precise lexical equivalent)

- *when he finishes his degree erm cos he's up here on a* sort of *second year degree the[i:] equivalent of about sixteen*

- 3. evidential (epistemic modal) *sort of* where the speaker avoids being precise due to, e.g., his lexical gap (Crystal – Davy's 1975 "memory loss)"

- 4. interpersonal (affective) *sort of* where *sort of* helps the speaker gain common ground with their interlocutor (Crystal – Davy's 1975 "keeping informal atmosphere with the speaker)

'sort of' - Polglish underuse

- 'Sort of' and 'kind of'

  (*kind of* as a possible culprit of the underuse)

- Repetition of the previous sequence

  (Polglish repair)

- Low incidence of VTs in speakers' L1 Polish ?

'sort of' – learner strategies

- *Kind of* - Polglish overuse of *kind of* prevails in the phraseological function, rather than the discursive one

- Repetition of the previous sequence

  ➢ *was sort of > I was I was*

this house was so old that I was actually I was I was scared *while walking on the stairs*

  ➢ *you can sort of > you can you can*

maybe you you can you can you know . uh you can you can buy some su= mm you can buy m= monthly supply of beer

L1 data preprocessing

- All turns used (character of the AOK corpus)
- Corpora used in their normalized word number

corpus	PLINDSEI	LOCNEC	AOK
tokens	76 064	75 589	75 853

Cluster, Concord WST tools
Functional analysis after Biber et al. (1999) and
De Cock (2004) categories – 1 extra category:
linking devices

'sort of' – Polish equivalents (1)

- Hypothesis: less VTs in Polish as L1 than in English as L1 = little VTs in English as L2

- List of Polish VTs meaning 'sort of" chosen from all Polish VTs

- vague use (V)separated from literal use (L)

- Comparison of frequencies with frequency dictionary of contemporary Polish

'sort of' – Polish equivalents (2)

- naprawdę
- jakiś (I)
- właściwie
- w zasadzie
- rzeczywiście
- w sumie
- (że) tak powiem
- raczej
- całkiem

- (tak) jakby
- niby
- dosyć
- tak trochę
- poniekąd
- (coś) w rodzaju
- (coś) w stylu
- pewien (I)
- nawet

- w (...) sensie
- to jest taki (I)
- ten (I)
- jak gdyby
- po prostu

'sort of' – Polish equivalents – V / all

- naprawdę 63/74
- właściwie 33/33
- w zasadzie 29/35
- rzeczywiście 22/27
- w sumie 22/24
- (że) tak powiem 21/21
- raczej12/16
- całkiem 13/13

- (tak) jakby 13/13
- dosyć 7/9
- tak trochę 5/5
- poniekąd 1/1
- (coś) w rodzaju 1/1
- (coś) w stylu 1/2

- nawet 9/80
- w (...) sensie 12/12
- to jest taki (l) 31/33
- pewien (l) 7/11
- niby 11/11
- po prostu 286

Some Polish examples V & L use

- V: No, to jest proste **w sumie**, nie? Dzieli ..

- L = altogether: Trzymaj te zdjęcia W sumie to byliśmy ... **w sumie** w czterech miejscach.

- V: i bierzesz coś tam i nikt cię **w zasadzie** nie obsługuje. Yhm Bo to jest, no ... cztery ścian

- L = as a rule:.natomiast „walkman"... **W zasadzie** to jest CASSETTE PLAYER...

- V: O tutaj to ja **nawet** chyba zaznaczyłam jakiś taki domek, który ewentualnie oni planują. ...

- L = even: Oni w Rumunii bardzo chętni są do brania **nawet** drobnych łapówek

Polish 'sort of' – frequency dictionary speech vs writing

- naprawdę 91v7
- właściwie 180v12
- w zasadzie --
- rzeczywiście 95v11
- w sumie --
- (że) tak powiem --
- raczej35v14
- całkiem 2v2

- to jest taki (l) 1305v195
- dosyć 34v2
- tak trochę 122v5

- poniekąd 0v1

- (coś) w rodzaju --
- (coś) w stylu --

- nawet 156v89

- (tak) jakby 60v8
- pewien (l) > (?)
- niby 15v1

- w (…) sensie --

'Sort of' – English synonyms

- kind of 264

- Some 147

- A 1832

- Generally 11

- in a way 10

- Really 532

- Impossibility to find all equivalents

- But: +/- only 300 Polish

Conclusions

- Characteristic features of learner language in terms of VTs:

    ➢ small amount of vagueness expressed

    ➢ substitution of one VT with another

    ➢ repetition of a sequence of words

- Features of Polish as L1

    ➢ abundance of various VTs but low percentage in general

    ➢ vague use over literal use

- Hypothesis confirmed but ...

References (1)

Aijmer, Karin. 2002. English discourse particles: Evidence from a corpus. Amsterdam: Benjamins.

Aijmer, Karin – Bengt Altenberg – Matts Johansson (eds.). 1996. Languages in contrast. Text-based cross-linguistic studies. Lund Studies in English 88. Lund: Lund University Press.

Biber, Douglas – Stig Johansson – Geoffrey Leech – Susan Conrad – Edward Finnegan (eds.). 1999. Longman grammar of spoken and written English. Harlow: Longman.

Crystal, David – Derek Davy. 1975. Advanced Conversational English. London: Longman.

De Cock, Sylvie. 1998. "Corpora of Learner Speech and Writing and ELT", in: A. Usoniene (ed.), 56-66.

References (2)

De Cock, Sylvie. 2004. "Preferred sequences of words in NS and NNS speech", Belgian Journal of English Language and Literatures (BELL), New Series 2: 225-246.

Duszak, Anna – Elżbieta Gajek – Urszula Okulska. 2006. Korpusy w angielsko-polskim językoznawstwie kontrastywnym.Kraków: universitas.

Granger, Sylviane. 1996. "From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora", in: Karin Aijmer – Bengt Altenberg – Matts Johansson (eds.), 37-51.

Grochowski, Maciej. 1986. Polskie partykuły. Wrocław: Ossolineum.

Kay, Paul. 1997. Words and the grammar of context. Chapter 6: The kind of/sort of construction. Stanford, Cal.: CSLI Publications. 145- 158.

References (3)

Kurcz, Ida – Andrzej Lewicki – Jadwiga Sambor – Krzysztof Szafran – Jerzy Woronczak. 1990. Słownik frekwencyjny polszczyzny współczesnej. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego w Krakowie.

Otwinowska-Kasztelanic, Agnieszka. 2000a. Korpus języka mówionego młodego pokolenia Polaków. Warszawa: Wydawnictwo Akademickie Dialog.

Otwinowska-Kasztelanic, Agnieszka. 2000b. A study of the lexico-semantic and grammatical influence of English on the Polish of the younger generation of Poles. Warszawa: Wydawnictwo Akademickie Dialog.

Perczyńska, Nina. 1975. Wybrane cechy składniowo-stylistyczne polszczyzny mówionej. Wrocław: Ossolineum.

References (4)

Pisarkowa, Krystyna. 1975. Składnia rozmowy telefonicznej. Wrocław: Ossolineum.

Scott, Mike. 1998. WordSmith Tools Manual, version 3.0. Oxford: Oxford University Press. (http://www.lexically.net/wordsmith/) (date of access: 26 Mar 2006)

Topolińska, Zuzanna (ed.). 1984. Gramatyka współczesnego języka polskiego: Składnia. [Grammar of contemporary Polish: Syntax.] Warszawa: PWN.

Usoniene, A. (ed.). 1998. Proceedings from the International Conference on Germanic and Baltic Linguistic Studies and Translation. University of Vilnius, 22-24 April 1998. Vilnius: Homo Liber.:

THANK
YOU