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

<table>
<thead>
<tr>
<th>corpus</th>
<th>PL</th>
<th>NS</th>
<th>FR</th>
</tr>
</thead>
<tbody>
<tr>
<td>tokens</td>
<td>90871</td>
<td>90945</td>
<td>90851</td>
</tr>
</tbody>
</table>
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.”
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:
<table>
<thead>
<tr>
<th></th>
<th>WORD</th>
<th>FREQ</th>
<th>RDB LST %</th>
<th>FREQ</th>
<th>NS50B TXT %</th>
<th>KEYNESS</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>FIRST OF ALL</td>
<td>15</td>
<td>0.02</td>
<td>0</td>
<td>24.8</td>
<td>0</td>
<td>0.000001</td>
</tr>
<tr>
<td>31</td>
<td>SHE DOESN'T LIKE</td>
<td>0</td>
<td>21</td>
<td>0.02</td>
<td>24.2</td>
<td>0</td>
<td>0.000001</td>
</tr>
<tr>
<td>32</td>
<td>BIT OF A</td>
<td>0</td>
<td>21</td>
<td>0.02</td>
<td>24.2</td>
<td>0</td>
<td>0.000001</td>
</tr>
<tr>
<td>33</td>
<td>I DON'T REALLY</td>
<td>0</td>
<td>21</td>
<td>0.02</td>
<td>24.2</td>
<td>0</td>
<td>0.000001</td>
</tr>
<tr>
<td>34</td>
<td>IT'S A BIT</td>
<td>0</td>
<td>22</td>
<td>0.02</td>
<td>25.3</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>35</td>
<td>I CAN'T REMEMBER</td>
<td>0</td>
<td>22</td>
<td>0.02</td>
<td>25.3</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>36</td>
<td>STUFF LIKE THAT</td>
<td>0</td>
<td>22</td>
<td>0.02</td>
<td>25.3</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>37</td>
<td>WAS A BIT</td>
<td>0</td>
<td>22</td>
<td>0.02</td>
<td>25.3</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>38</td>
<td>THOUGHT IT WAS</td>
<td>0</td>
<td>23</td>
<td>0.02</td>
<td>26.5</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>39</td>
<td>WANTED TO DO</td>
<td>0</td>
<td>23</td>
<td>0.02</td>
<td>26.5</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>40</td>
<td>I MEAN IT</td>
<td>0</td>
<td>23</td>
<td>0.02</td>
<td>26.5</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>41</td>
<td>IN THE MORNING</td>
<td>0</td>
<td>23</td>
<td>0.02</td>
<td>26.5</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>42</td>
<td>TO SORT OF</td>
<td>0</td>
<td>23</td>
<td>0.02</td>
<td>26.5</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>43</td>
<td>YEAH I</td>
<td>0</td>
<td>24</td>
<td>0.02</td>
<td>27.5</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>44</td>
<td>I MEAN IT'S</td>
<td>0</td>
<td>24</td>
<td>0.02</td>
<td>27.5</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>45</td>
<td>END OF</td>
<td>0</td>
<td>24</td>
<td>0.02</td>
<td>27.5</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>46</td>
<td>LIKE TO GO</td>
<td>0</td>
<td>24</td>
<td>0.02</td>
<td>27.5</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>47</td>
<td>A LOT MORE</td>
<td>0</td>
<td>24</td>
<td>0.02</td>
<td>27.5</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>48</td>
<td>THAT SORT OF</td>
<td>0</td>
<td>24</td>
<td>0.02</td>
<td>27.5</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>49</td>
<td>WAS REALLY GOOD</td>
<td>0</td>
<td>24</td>
<td>0.02</td>
<td>27.5</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>50</td>
<td>ALL HER FRIENDS</td>
<td>0</td>
<td>25</td>
<td>0.02</td>
<td>28.8</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>51</td>
<td>A SORT OF</td>
<td>0</td>
<td>25</td>
<td>0.02</td>
<td>28.8</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>52</td>
<td>THINGS LIKE THAT</td>
<td>8</td>
<td>56</td>
<td>0.05</td>
<td>29.5</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>53</td>
<td>SO THAT WAS</td>
<td>0</td>
<td>28</td>
<td>0.02</td>
<td>32.2</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>54</td>
<td>BUT I MEAN</td>
<td>0</td>
<td>31</td>
<td>0.03</td>
<td>35.7</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>55</td>
<td>AND THINGS LIKE</td>
<td>0</td>
<td>31</td>
<td>0.03</td>
<td>35.7</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>56</td>
<td>YEAH THAT'S RIGHT</td>
<td>0</td>
<td>32</td>
<td>0.03</td>
<td>36.8</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>57</td>
<td>I'D LIKE TO</td>
<td>6</td>
<td>62</td>
<td>0.06</td>
<td>40.7</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>58</td>
<td>THAT'S RIGHT YEAH</td>
<td>0</td>
<td>37</td>
<td>0.03</td>
<td>42.6</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>59</td>
<td>YEAH YEAH</td>
<td>0</td>
<td>40</td>
<td>0.03</td>
<td>45.0</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>60</td>
<td>SORT OF THING</td>
<td>0</td>
<td>44</td>
<td>0.04</td>
<td>50.6</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>61</td>
<td>I MEAN I</td>
<td>8</td>
<td>83</td>
<td>0.07</td>
<td>54.6</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>62</td>
<td>AT THE MOMENT</td>
<td>0</td>
<td>49</td>
<td>0.04</td>
<td>56.4</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>63</td>
<td>A BIT OF</td>
<td>0</td>
<td>49</td>
<td>0.04</td>
<td>56.4</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>64</td>
<td>SORT OF LIKE</td>
<td>0</td>
<td>50</td>
<td>0.04</td>
<td>57.6</td>
<td>0</td>
<td>0.000000</td>
</tr>
<tr>
<td>65</td>
<td>YEAH YEAH</td>
<td>0</td>
<td>67</td>
<td>0.06</td>
<td>77.1</td>
<td>0</td>
<td>0.000000</td>
</tr>
</tbody>
</table>
'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)

3. 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

<table>
<thead>
<tr>
<th>corpus</th>
<th>PLINDSEI</th>
<th>LOCNEC</th>
<th>AOK</th>
</tr>
</thead>
<tbody>
<tr>
<td>tokens</td>
<td>76 064</td>
<td>75 589</td>
<td>75 853</td>
</tr>
</tbody>
</table>

_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 (l)
- 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 (l)
- nawet
- w (...) sensie
- to jest taki (l)
- ten (l)
- 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
- raczej 12/16
- całkiem 13/13
- (tak) jakby 13/13
- dosyć 7/9
- tak trochę 5/5
- nawet 9/80
- w (...) sensie 12/12
- to jest taki (l) 31/33
- pewien (l) 7/11
- niby 11/11
- (coś) w rodzaju 1/1
- (coś) w stylu 1/2
- 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 --
- raczej 35v14
- całkiem 2v2
- to jest taki (l) 1305v195
- dosyć 34v2
- tak trochę 122v5
- nawet 156v89
- (tak) jakby 60v8
- pewien (l) > (?)
- niby 15v1
- (coś) w rodzaju --
- (coś) w stylu --
- 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)


References (3)


References (4)


THANK YOU