Rethinking the German three-way system of spatial demonstrative adverbs: evidence from electronic corpora

Johnathan L. W. Gajdos
johnathan-gajdos@uiowa.edu
Department of German
University of Iowa

October 9, 2009
AACL 2009, University of Alberta
Context

• Demonstrative adverbs: speakers refer to objects in the context of the location relative to the speaker-hearer interaction (the local anchor and the location identified by the adverb are both relevant).

• English: *here, there*

• German: *hier ‘here’, dort ‘there’, da ‘here/there’*
English

• Deictic choice that depends on speaker location
• Can have anaphoric use, but generally deictic function
• The speaker must use here if (s)he is in London at the time of utterance and there if (s)he is not. There is a deictic relationship between source and target (Lyons 1977):
  a. I was born in London and have lived here all my life.
  b. I was born in London and have lived there all my life.
German

- Tripartite system in German: rough equivalents *hier* ‘here’ and *dort* ‘there’, but also a third choice—
- Have deictic function, but can also function anaphorically
- Karl Valentin ist in München geboren, und er ist hier/dort/da auch aufgewachsen. (Lenz 2001) ‘Karl Valentin was born in Munich, and he also grew up here/there/here-there.’

CLAIM: *da* as neutral form—no information provided about location of speaker
Ich bin vor vier Jahren von Düsseldorf nach Nijmegen gegangen und *hier/dort/da* will ich vorläufig bleiben. (Ehrich 1982)

‘Four years ago I moved from Düsseldorf to Nijmegen, and here/there/here-there I want to stay for some time.’

CLAIM: *hier* can only be used when speaker is in Nijmegen and *dort* only if not in Nijmigen; is *da* more versatile?
Defining *da*

**Wahrig (2005)**

- “*hier, an dieser Stelle; dort, an jener Stelle*”
  ‘here, at this place; there, at that place’

- a single numbered definition in the entry for *da*
Defining *da*

Duden Universalwörterbuch (2006)

- Two definitions of *da*—*dort*-focused and *hier*-focused
- *dort*-centric (10 examples total)
  - *da* ist die Haltestelle
    ‘there is the (bus) stop’
  - es muss noch Brot *da*
    ‘there must still be bread (Ø/here/there)’
  - diese Dinge sind dazu *da*, dass man sie benutzt
    ‘these things are here/there to be used’
  - er wohnt *da*
    ‘he lives there/here’
Defining *da*

Duden Universalwörterbuch (2006)

- Two definitions of *da*—*dort*-focused and *hier*-focused
- *Hier*-centric (3 examples total)
  - *da* sind wir
    ‘here/there we are’
  - ist *da* jemand
    ‘is someone there/here?’
  - *da*, nimm das Geld
    ‘here, take the money’
An empirical approach—
corpus data and survey/judgment data

Research questions
• What can account for the distribution and use of the members of the German three-way demonstrative adverb system of spatial deixis?
  – Are corpus data consistent with native speaker survey data?
  – What is the meaning that should be assigned to *da*?
  – What accounts for the selection, preference, or dispreference for *da* in an utterance?
Research questions, continued

• How do L1 child learners of German make use of these spatial expressions, and how is it the same or different from adult native and non-native speakers of German?
Further research questions

– How is this system influenced by semantic-pragmatic conditions?
– Is there evidence for variation based on type of discourse (e.g., spoken v. written)?
– Is there evidence for regional variation?
Research questions from related research (survey data/grammaticality judgments)

• How do L2 speakers of German interpret and use the members of this demonstrative adverb system?
  – Is there a learner gap? If so, at which levels of learning?
  – Does there appear to be L1 transfer for English native speakers learning German?
  – Do highly advanced non-native-speakers of German exhibit learner-like, native-like, or distinct patterns?
# Part I: Corpus analysis

## Overview of corpora used (Figure 1)

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Number of words</th>
<th>Time period</th>
<th>Number of works</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Juilland-D</td>
<td>500,000</td>
<td>1920-1939</td>
<td>398</td>
<td>Distribution by number of words: Drama: 20% Novels, short stories: 20% Essays: 20% Journalism: 20% Scientific writing: 20%</td>
</tr>
<tr>
<td>DWDS Core Corpus (Kerncorpus)</td>
<td>122,816,010 (2,224,542 distinct words)</td>
<td>1900-2000</td>
<td>79,830</td>
<td>Distribution by number of words: Fiction: 26% Technical writing: 22% Scientific writing: 24% Newspapers: 28% Distribution by number of texts: Literature: 26% Journalism: 27% Technical writing: 20% Spoken transcription: 5%</td>
</tr>
<tr>
<td>DDR corpus</td>
<td>9,000,000</td>
<td>1949-1990</td>
<td>1150</td>
<td>Currently, public documents</td>
</tr>
<tr>
<td>Die Zeit corpus</td>
<td>106,000,000</td>
<td>1946-2008</td>
<td>493,380</td>
<td>A single German-language newspaper</td>
</tr>
</tbody>
</table>
Part I: General text corpus analysis
Basic corpus search patterns for corpus sentences

<table>
<thead>
<tr>
<th></th>
<th>hier</th>
<th>Da</th>
<th>dort</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Search restricted to sentence boundaries—

Where do these words co-occur?
Can indicated locations be determined?

Figure 2
Percent of corpus sentences containing/excluding specified words (Figure 3)

<table>
<thead>
<tr>
<th>Words present in single sentence</th>
<th>Words not present in single sentence</th>
<th>Juillard-D (32,846)</th>
<th>DDR (396,047)</th>
<th>Spoken (1,39,633)</th>
<th>Die Zeit (17.7M)</th>
<th>DWDS Core: Technical writing</th>
<th>DWDS Core: Scientific texts</th>
<th>DWDS Core: Fiction (1.8M)</th>
<th>DWDS Core: Newspapers (1.4M)</th>
<th>DWDS Core: All genres (6.0M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>da</td>
<td>hier, dort</td>
<td>1.796</td>
<td>1.061</td>
<td>3.441</td>
<td>1.823</td>
<td>0.783</td>
<td>0.448</td>
<td>2.169</td>
<td>0.825</td>
<td>1.135</td>
</tr>
<tr>
<td>da</td>
<td>hier, dort</td>
<td>1.681</td>
<td>0.997</td>
<td>3.214</td>
<td>1.745</td>
<td>0.735</td>
<td>0.410</td>
<td>2.046</td>
<td>0.770</td>
<td>1.065</td>
</tr>
<tr>
<td>da, dort, hier</td>
<td>hier, dort</td>
<td>0.003</td>
<td>0.002</td>
<td>0.007</td>
<td>0.002</td>
<td>0.002</td>
<td>0.001</td>
<td>0.004</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>dort</td>
<td>da, hier</td>
<td>0.496</td>
<td>0.527</td>
<td>0.851</td>
<td>0.985</td>
<td>0.676</td>
<td>0.467</td>
<td>0.753</td>
<td>0.606</td>
<td>0.634</td>
</tr>
<tr>
<td>dort</td>
<td>da, hier</td>
<td>0.411</td>
<td>0.475</td>
<td>0.731</td>
<td>0.903</td>
<td>0.636</td>
<td>0.403</td>
<td>0.659</td>
<td>0.565</td>
<td>0.573</td>
</tr>
<tr>
<td>hier</td>
<td>da, dort</td>
<td>1.611</td>
<td>1.314</td>
<td>3.301</td>
<td>1.851</td>
<td>1.407</td>
<td>1.725</td>
<td>1.693</td>
<td>1.234</td>
<td>1.527</td>
</tr>
<tr>
<td>hier</td>
<td>da, dort</td>
<td>1.498</td>
<td>1.233</td>
<td>3.085</td>
<td>1.736</td>
<td>1.348</td>
<td>1.649</td>
<td>1.563</td>
<td>1.169</td>
<td>1.441</td>
</tr>
<tr>
<td>da, dort</td>
<td>hier</td>
<td>0.043</td>
<td>0.017</td>
<td>0.062</td>
<td>0.021</td>
<td>0.013</td>
<td>0.012</td>
<td>0.041</td>
<td>0.015</td>
<td>0.022</td>
</tr>
<tr>
<td>dort, hier</td>
<td>da</td>
<td>0.040</td>
<td>0.034</td>
<td>0.051</td>
<td>0.059</td>
<td>0.025</td>
<td>0.050</td>
<td>0.048</td>
<td>0.025</td>
<td>0.038</td>
</tr>
<tr>
<td>hier, da</td>
<td>dort</td>
<td>0.070</td>
<td>0.046</td>
<td>0.158</td>
<td>0.054</td>
<td>0.033</td>
<td>0.025</td>
<td>0.077</td>
<td>0.040</td>
<td>0.046</td>
</tr>
</tbody>
</table>
% of sentences containing *da*, *dort*, and *hier*, by corpus (Figure 4)

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
<th>h</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td>da</td>
<td>1.80</td>
<td>1.06</td>
<td>3.44</td>
<td>1.82</td>
<td>0.78</td>
<td>0.45</td>
<td>2.17</td>
<td>0.83</td>
<td>1.14</td>
</tr>
<tr>
<td>dort</td>
<td>0.50</td>
<td>0.53</td>
<td>0.85</td>
<td>0.99</td>
<td>0.68</td>
<td>0.47</td>
<td>0.75</td>
<td>0.61</td>
<td>0.63</td>
</tr>
<tr>
<td>hier</td>
<td>1.61</td>
<td>1.31</td>
<td>3.30</td>
<td>1.85</td>
<td>1.41</td>
<td>1.73</td>
<td>1.69</td>
<td>1.23</td>
<td>1.53</td>
</tr>
</tbody>
</table>

Figure 4
% of sentences with exactly two of the three adverbs (Figure 5)

Sentences containing the third word are excluded

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
<th>h</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td>da, dort</td>
<td>0.043</td>
<td>0.017</td>
<td>0.062</td>
<td>0.021</td>
<td>0.013</td>
<td>0.012</td>
<td>0.041</td>
<td>0.015</td>
<td>0.022</td>
</tr>
<tr>
<td>dort, hier</td>
<td>0.04</td>
<td>0.034</td>
<td>0.051</td>
<td>0.059</td>
<td>0.025</td>
<td>0.05</td>
<td>0.048</td>
<td>0.025</td>
<td>0.038</td>
</tr>
<tr>
<td>hier, da</td>
<td>0.07</td>
<td>0.046</td>
<td>0.158</td>
<td>0.054</td>
<td>0.033</td>
<td>0.025</td>
<td>0.077</td>
<td>0.04</td>
<td>0.046</td>
</tr>
</tbody>
</table>

- a Juillard-D
- b DDR
- c Spoken
- d Die Zeit
- e Technical writing (core)
- f Scientific texts (core)
- g Fiction (core)
- h newspapers (core)
- i Entire DWDS core
% of sentences containing only one of the three adverbs

- only hier
- only dort
- only da

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
<th>h</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td>only hier</td>
<td>1.498</td>
<td>1.233</td>
<td>3.085</td>
<td>1.736</td>
<td>1.348</td>
<td>1.649</td>
<td>1.563</td>
<td>1.169</td>
<td>1.441</td>
</tr>
<tr>
<td>only dort</td>
<td>0.411</td>
<td>0.475</td>
<td>0.731</td>
<td>0.903</td>
<td>0.636</td>
<td>0.403</td>
<td>0.659</td>
<td>0.565</td>
<td>0.573</td>
</tr>
<tr>
<td>only da</td>
<td>1.681</td>
<td>0.997</td>
<td>3.214</td>
<td>1.745</td>
<td>0.735</td>
<td>0.41</td>
<td>2.046</td>
<td>0.77</td>
<td>1.065</td>
</tr>
</tbody>
</table>

Figure 6
Ratios—sentences containing/excluding specified words:
sentences containing *da* (Figure 7)

<table>
<thead>
<tr>
<th>Search terms</th>
<th>Search terms</th>
<th>range</th>
<th>median</th>
<th>mean</th>
<th>Juilland-D</th>
<th>DDR</th>
<th>Spoken</th>
<th>Die Zeit (newspaper)</th>
<th>DWDS Core: Technical writing</th>
<th>DWDS Core: Scientific texts</th>
<th>DWDS Core: Fiction</th>
<th>DWDS Core: newspapers</th>
<th>DWDS Core: All genres</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>da</em></td>
<td>hier, dort</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td><em>da</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>da</em></td>
<td>hier, dort</td>
<td>0.05</td>
<td>1.07</td>
<td>1.07</td>
<td>1.06</td>
<td>1.07</td>
<td>1.04</td>
<td>1.06</td>
<td>1.09</td>
<td>1.06</td>
<td>1.07</td>
<td>1.07</td>
<td>1.07</td>
</tr>
<tr>
<td><em>da</em>, dort, hier</td>
<td></td>
<td>608.4</td>
<td>545.7</td>
<td>606.7</td>
<td>590.0</td>
<td>700.5</td>
<td>480.5</td>
<td>808.2</td>
<td>493.5</td>
<td>344.6</td>
<td>544.0</td>
<td>953.0</td>
<td>545.7</td>
</tr>
<tr>
<td>dort</td>
<td></td>
<td>3.08</td>
<td>1.85</td>
<td>2.19</td>
<td>3.62</td>
<td>2.01</td>
<td>4.04</td>
<td>1.85</td>
<td>1.16</td>
<td>0.96</td>
<td>2.88</td>
<td>1.36</td>
<td>1.79</td>
</tr>
<tr>
<td>dort</td>
<td><em>da</em>, hier</td>
<td>3.60</td>
<td>2.02</td>
<td>2.49</td>
<td>4.37</td>
<td>2.23</td>
<td>4.71</td>
<td>2.02</td>
<td>1.23</td>
<td>1.11</td>
<td>3.29</td>
<td>1.46</td>
<td>1.98</td>
</tr>
<tr>
<td>hier</td>
<td></td>
<td>1.02</td>
<td>0.81</td>
<td>0.83</td>
<td>1.12</td>
<td>0.81</td>
<td>1.04</td>
<td>0.98</td>
<td>0.56</td>
<td>0.26</td>
<td>1.28</td>
<td>0.67</td>
<td>0.74</td>
</tr>
<tr>
<td>hier</td>
<td><em>da</em>, dort</td>
<td>1.12</td>
<td>0.86</td>
<td>0.89</td>
<td>1.20</td>
<td>0.86</td>
<td>1.12</td>
<td>1.05</td>
<td>0.58</td>
<td>0.27</td>
<td>1.39</td>
<td>0.71</td>
<td>0.79</td>
</tr>
<tr>
<td><em>da</em> dort</td>
<td>hier</td>
<td>47.72</td>
<td>55.87</td>
<td>55.82</td>
<td>42.14</td>
<td>62.73</td>
<td>55.87</td>
<td>84.87</td>
<td>59.33</td>
<td>37.15</td>
<td>52.29</td>
<td>56.06</td>
<td>51.93</td>
</tr>
<tr>
<td><em>dort</em>, hier</td>
<td><em>da</em></td>
<td>58.74</td>
<td>31.37</td>
<td>35.99</td>
<td>45.38</td>
<td>31.37</td>
<td>67.68</td>
<td>31.01</td>
<td>31.29</td>
<td>8.94</td>
<td>45.23</td>
<td>32.96</td>
<td>30.02</td>
</tr>
<tr>
<td><em>hier</em>, <em>da</em></td>
<td><em>dort</em></td>
<td>15.74</td>
<td>23.91</td>
<td>24.39</td>
<td>25.65</td>
<td>23.09</td>
<td>21.74</td>
<td>33.62</td>
<td>23.91</td>
<td>17.88</td>
<td>28.08</td>
<td>20.87</td>
<td>24.66</td>
</tr>
</tbody>
</table>
Corpus examples

All 3 co-occur; refer to 3 locations

   ‘But everything in connection with Pippi has changed so much [that] today she appears here as a doll and there as a picture and there as a movie character.’

   ‘The book goes against those who prefer to nibble from here, there, and there.’

   ‘When that begins, everyone whines: not here/there, not there, and not here.’

   ‘There lay the waterfowl, here the horse, and there the small lion man’
And all that they found there: cafés and jazz bars; antique stores and well-stocked stores, which, eager for foreign currency, stayed open until late in the evening. But who was to spend money there, if they all stayed here (in their hotels)?
Corpus examples

4 occurrences in a sentence; 4 locations

- Da$_{\text{Loc1}}$ und dort$_{\text{Loc2}}$ gibt es ein paar Abgründe, hier$_{\text{Loc3}}$ und da$_{\text{Loc4}}$ ein paar Leute, die außer dem Internet nichts haben. (Die Zeit, 9/2005)

‘Here and there there are a few abysses; here and there a few people, who except for the Internet, have nothing.’

- Das Kind stellt Gefäße unter die Rinnsale, aus Plastik, aus Glas, aus Metall, es trommelt hier$_{\text{Loc1}}$, und da$_{\text{Loc2}}$ klatscht es, dort$_{\text{Loc3}}$ prasselt es, und hier$_{\text{Loc4}}$ zirpt es, die Wassermusik füllt das Haus. (Die Zeit, 7/2005)

‘The child places containers under the gutters—plastic, glass, and metal—it drums here, there it claps, it crackles there, and here it chirps; the water music fills the house.’
Part 2: Child-language corpus analysis (CHILDES)
Overview of corpora used

- CHILDES
- 6 German-language corpora
  - German language—83 children total
  - Age range: 0.15-12.23 yrs.
  - Median age: 3.05 yrs.
- Also includes speech of parents or other adults
Part 2: Child-language corpus analysis (CHILDES)
Specific corpora used

- Caroline; ages 0;10 – 4;3; 1 child
- Miller; ages 0;10-4;0; 3 children
- Rigole; ages 0-7; 4 children
- Szagun; ages 1;6-3;48 children
- Wagner; ages 1;5–14;13 children
- Weissenborn; ages 7–11; 14 children
Use of *da*, *dort*, and *hier*, by age of child

Figure 8
Overview of average tokens/utterance in six L1 child language corpora (CHILDES)

Figure 9

- da/utt
- hier/utt
- dort/utt
Overview of average tokens/utterance in six L1 child language corpora (CHILDES)--non-Child participants

Figure 10
Conclusion

- a simple account of *hier/da/dort* use is problematic; previous accounts seem inadequate
- *da* should not be viewed as being devoid of semantic information
- *da* has neutral characteristics, but also *dort*-replacement tendencies
Further research/analyses

- MLU variation with L1
- Individual children in CHILDES corpora
- Further corpus searches; hand-coding of results
- Additional spoken-language corpora
- Incorporation of L2 data in the context of NS use
- L1 acquisition/early child L1 use
- Genre differences; speech v writing
References


Lenz, Friedrich. 2001. Here is hier, and there is dort, but where is da?: Contrastive reflections on English and German demonstrative adverbs. Arbeiten aus Anglistik und Amerikanistik 26. 39-51.


