Creating an Annotated Corpus for Generating Walking Directions

Stephanie Schuldes† **Michael Roth**‡ **Anette Frank**‡ **Michael Strube**†

†EML Research gGmbH
Heidelberg, Germany
http://www.eml-research.de/nlp

‡Department of Computational Linguistics
University of Heidelberg, Germany
http://www.cl.uni-heidelberg.de

Abstract

This work describes first steps towards building a system that synchronously generates multimodal (textual and visual) route directions for pedestrians. We pursue a corpus-based approach for building a generation model that produces natural instructions in multiple languages. We conducted an empirical study to collect verbal route directions, and annotated the acquired texts on different levels. Here we describe the experimental setting and an analysis of the collected data.

1 Introduction

Route directions guide a person unfamiliar with the environment to their designated goal. We plan to generate route instructions that are similar to those given by humans by referring to landmarks and by structuring the route in a way that it is easy to memorize (Denis, 1997).

We develop a system for synchronously generating natural language route directions and 3D scenes of a route. The core of the architecture is a unified representation providing information for both verbal and graphical output. The direct correspondence between linguistic references and shown objects facilitates the identification of the visual scene in the real world and the choice of the correct action while following the route. To create a reusable system that is adaptable to different navigational domains and languages, we use machine learning techniques to build a statistical generation model from annotated corpora. We report on an empirical study to collect human-produced walking directions to be used for statistical generation from underlying semantic structures. While our scenario is ultimately multilingual, here we give an analysis of the German dataset.

2 Related Work

The task of analyzing and generating cognitively adequate route instructions has been addressed by a number of authors (Taylor & Tversky, 1996; Tappe, 2000; Habel, 2003; Richter, 2008; Viethen & Dale, 2008; Kelleher & Costello, 2009). Marciniak & Strube (2005) showed that a system for generating route directions can be successfully trained on a small set of 75 route direction texts (8418 tokens). In their approach directions are represented in a graph, which encodes information on various conceptual levels. While their approach is restricted to reproducing directions for the learned graphs, we will generate directions for a wide range of possible routes. Dale et al. (2005) developed a system that takes GIS data as input and uses a pipeline architecture to generate verbal route directions. In contrast to their approach, our approach will be based on an integrated architecture allowing for more interaction between the different stages of generation. The idea of combining verbal directions with scenes from a virtual 3D environment has recently lead to a new framework for evaluating NLG systems: The Challenge on Generating Instructions in Virtual Environments (GIVE) (Byron et al., 2009) is planned to become a regular event for the NLG community.

3 Corpus Acquisition

For collecting naturally produced route instructions, we conducted a study with 29 native speakers of German (66% female and 33% male). The participants in our study were students from various fields aged between 20 and 34 years. We designed two different settings: one *on-site setting*, in which participants walked around in a real world situation (specifically our university campus), and one *desk-based setting*, in which they interacted with a web application. The former was further divided into indoor and outdoor routes,

Proceedings of the 2009 Workshop on Language Generation and Summarisation, ACL-IJCNLP 2009, pages 72–76,
Suntec, Singapore, 6 August 2009.

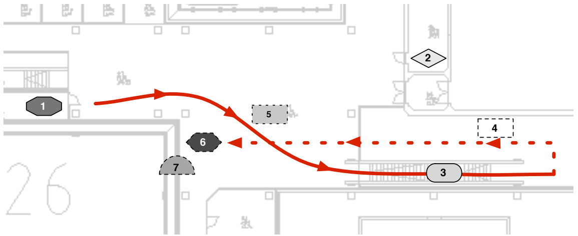

Figure 1: Example route from the indoor setting (first task), leading from a room with photocopiers (1) across an open space and downstairs (3) to a students' union room (6), passing an information board (4) and a coffee machine (5). A lecture room (2) and a glass wall (7) are visible from the route.

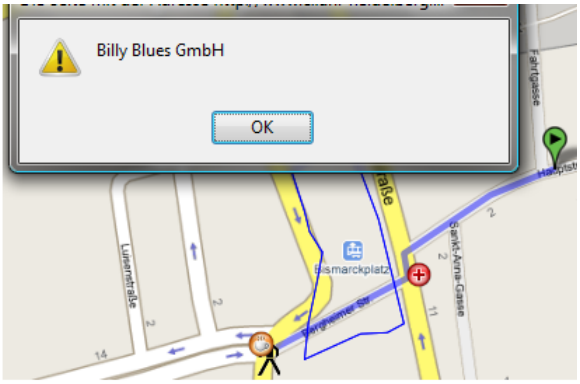

Figure 2: Web application used in the second task. Landmarks were introduced successively via pop-ups as the animated walker encountered them.

while the latter was restricted to an outdoor scenario. This design enables us to study possible differences and commonalities between linguistic realizations obtained for different environments as well as different presentation modes.

For both scenarios, the task was to give written directions to a person unfamiliar with the area as to how to get to the destination the participants just reached, taking the same route. First, participants were led along a route to a given destination point (on-site). Each participant was asked to give directions for two routes inside buildings of the university campus (e.g. from an office to a seminar room, cf. Figure 1), and one outside route (e.g. from the building entrance to a bus stop).

Second, participants were shown a web application that guided them along a route by means of a 2D animation (desk-based). Subjects were allowed to use all information displayed by the web application: named places, buildings, street and bridge names, etc. (cf. Figure 2).

Setting	GM	CI	CO	Total
physical routes	9	6	3	18
directions	59	58	28	145
tokens	5353	4119	2674	12146
tokens/dir. (∅)	91	71	96	

Table 1: Number of routes, directions, and tokens for the different settings. GM = Google Maps, CI = Campus Indoor, CO = Campus Outdoor.

4 Corpus Annotation

The acquired texts were processed in several steps. To ensure that all route directions consist of syntactically and semantically correct sentences, we manually corrected spelling mistakes, omissions resulting in grammatical errors, and removed elliptical and unclear directions.

The preprocessed texts were annotated on the following three levels:

pos_lemma – part-of-speech and lemma

syn_dep – dependency relations

sem_frame – frames and semantic roles

For the *pos_lemma* and *syn_dep* levels, we used *TreeTagger* (Schmid, 1997) and XLE (Maxwell & Kaplan, 1993). The corpus was parsed with the German ParGram LFG grammar (Forst, 2007). The outputs were corrected manually by two annotators. On the *sem_frame* level annotation was carried out using the annotation tool SALTO (Burchardt et al., 2006) and following the definiton of the FrameNet frames SELF_MOTION, PERCEPTION, BEING_LOCATED and LOCATIVE_RELATION (Baker et al., 1998). In terms of accuracy for unlabeled/labeled relations, the annotation agreement was 78.88%/65.17% on the *syn_dep* level and 79.27%/68.39% for frames and semantic roles.

5 Data Analysis

5.1 Corpus Statistics

We examined word frequencies with respect to the experimental settings in order to determine similarities and dissimilarities in lexical choice. Table 2 shows the three most frequent verbs and nouns found in each corpus part.

The data reveals that the most frequent verbs are typical among all settings. However, we found a number of lower-frequency verbs that are rather

Top verbs (Campus)	GM	CI	CO
gehen 'to walk'	11%	18%	14%
sein 'to be'	3.9%	8.2%	6.6%
stehen 'to stand'	**0.0%**	6.3%	5.3%
Top verbs (GM)	**GM**	**CI**	**CO**
folgen 'to follow'	12%	2.9%	2.6%
gehen 'to walk'	11%	18%	14%
abbiegen 'to turn into'	9.0%	3.8%	8.9%
Top nouns (Campus)	**GM**	**CI**	**CO**
Tür 'door'	**0.0%**	12%	0.9%
Treppe 'stairs'	**0.0%**	8.3%	**0.0%**
Gang 'hallway'	**0.0%**	6.6%	**0.0%**
Top nouns (GM)	**GM**	**CI**	**CO**
...straße '... Street'	28%	**0.0%**	2.2%
Richtung 'direction'	3.5%	2.8%	2.6%
...platz '... Square'	3.4%	**0.0%**	6.1%

Table 2: Relative frequency of the three most common verbs and nouns in both studies

scenario-specific. In many cases, the occurrence or absence of a verb can be attributed to a verb's selectional restrictions. For example, some of the verbs describing movements along streets (e.g. *folgen* 'to follow', *abbiegen* 'to turn into') do not occur within the indoor corpus whereas verbs describing "3D movements" (e.g. *durchqueren* 'to walk through', *hinuntergehen* 'to walk down') are not mentioned with the Google Maps setting.

The most frequent nouns significantly differ between the indoor and outdoor settings. This correlation does not come as a surprise, as most of the mentioned objects cannot be found in all scenarios. On the other hand, nouns that are common to both indoor and outdoor scenarios can be divided into two categories: Nouns denoting (1) objects that appear in both scenarios (e.g. *Gebäude* 'building') and (2) abstract concepts typical for route directions in general, e.g. *Richtung* 'direction', *Nummer* 'number', *Ziel* 'goal', and *Startpunkt* 'starting point'.

5.2 Landmark Alignment

Landmark alignment serves the purpose of detecting objects that are most frequently mentioned across directions, and how the same object is referred to differently. We created a graph-based representation of the landmarks mentioned in each route instruction (*single route representation, SRR*) for use in two types of alignment. Figure 3 shows an example from the indoor study. First, we created a combined graph for each physical route by merging the respective SRRs, taking into account several criteria:

String matching of landmark names;

Semantic similarity using *GermaNet* (Lemnitzer & Kunze, 2002), a lexical-semantic network for German similar to WordNet;

Frequency of references across all directions;

Spatio-temporal proximity of references to the same object;

Number of landmarks mentioned in a single direction (i.e. length of the SRR).

The combined graphs show that there are strong correspondences between the directions for the same route. We also found that, in the campus settings, there was a small number of frequently used general objects and a large number of less frequently used specific objects. This facilitates merging and shows the importance of the objects for people's orientation, and at the same time supports our claim that other modalities are needed to disambiguate references during navigation. For generating informative referential expressions, the combined graph needs to be refined so that object properties are represented (Krahmer et al., 2003).

Second, we aligned the SRRs with the physical route graph. Comparing the landmarks mentioned in the campus settings revealed that, in 97.8% of the cases, people adhere to the sequence in which objects are encountered. Reversed order was only found in special cases like distant objects.

5.3 Discourse Phenomena

We analyzed the use of anaphora, the temporal order of instructions, and occurrences of prototypical event chains in the collected texts in order to identify coherence-inducing elements.

Spatio-temporal adverbials: Most anaphors mention intermediate goals on the route in order to refer to the starting point of a new action (e.g. *da/hier* 'here', *dort* 'there'). This finding goes hand in hand with the observation that the collected route directions are typically structured in a linear temporal order (cf. Table 3) as for example indicated by the use of **adverbs indicating temporal succession** (e.g. *jetzt* 'now', *dann* 'then' and *danach* 'afterwards') and conjunctions (e.g. *bis* 'until', *wenn* 'when'). Interestingly, a reversed order can be found in a few cases, where

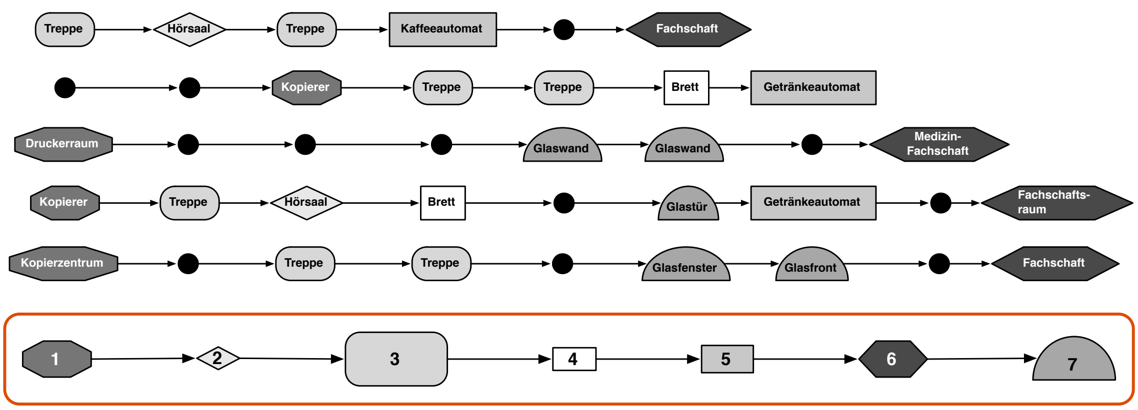

Figure 3: Each line shows one SRR for the route in Figure 1. Correspondences are indicated by identical node shapes, black dots substitute non-matched tokens. The bottom graph shows the physical route seen as sequence of landmarks. Node size reflects the importance of the referred object as conveyed by SRRs.

Adverbs $>_t$	**GM**	**CI**	**CO**
dann 'then'	55	43	30
jetzt 'now'	4	7	5
danach 'afterwards'	12	5	3
Adverbs $<_t$	**GM**	**CI**	**CO**
vorher 'beforehand'	0	1	0
davor 'before'	1	0	2

Table 3: Frequencies of temporal adverbs indicating linear ($>_t$) and reversed linear order ($<_t$)

the following action or situation is not supposed to take place (e.g. *Gehen Sie vorher rechts* 'beforehand turn right').

Backward-looking event anaphors and references to result states: We also found explicit references to past events (e.g. *Nach dem Durchqueren* 'after traversing') and result states of events, e.g. the adverbial phrase *unten angekommen* (here: 'downstairs') was frequently used following an instruction to 'walk downstairs'.

6 Conclusions and Future Work

The lexical corpus analysis confirms our hypothesis that there are strong commonalities in lexical choice for directions that persist across scenarios and presentation modes, with a small number of focused differences, and obvious domain-dependent lexical differences regarding the nature of objects in the respective scenarios. While our current corpus data is rather broad, environment-specific data can be extended quickly by setting up web studies using 2D and 3D environments.

The alignment of the physical routes and verbal instructions shows a clear tendency that linear route structure is observed in verbal realization, with only few exceptions. Since temporal order is observed by default, temporal annotation can be restricted to capture exceptional orderings, which are recoverable from linguistic cues. The study of discourse coherence effects yielded a number of elements that will be given special attention in the surface generation model. We observed a variety of coherence-inducing elements that are generic in nature and thus seem well-suited for a corpus-based generation model. As other languages are known to exhibit differences in verbal realization of directions (von Stutterheim et al., 2002), we have to extend our data collection in order to generate systematic linguistic variations from a single underlying semantic structure for all languages.

The linguistic annotation levels of frames and roles, syntactic dependencies, and basic word categories have been tested successfully with a similar corpus (Roth & Frank, 2009). The next steps will consist in the alignment of physical routes and landmarks with semantic representations in an integrated generation architecture.

Acknowledgements: This work is supported by the DFG-financed innovation fund FRONTIER as part of the Excellence Initiative at Heidelberg University (ZUK 49/1) and partially funded by the Klaus Tschira Foundation, Heidelberg, Germany. We thank the participants in our study, our annotators Tim Krones and Anna Schmidt, and student assistants Jonathan Geiger and Carina Silberer.

References

Baker, Collin F., Charles J. Fillmore & John B. Lowe (1998). The Berkeley FrameNet Project. In *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics,* Montréal, Québec, Canada, 10–14 August 1998, pp. 86–90.

Burchardt, Aljoscha, Katrin Erk, Anette Frank, Andrea Kowalski & Sebastian Pado (2006). SALTO: A versatile multi-level annotation tool. In *Proceedings of the 5th International Conference on Language Resources and Evaluation,* Genoa, Italy, 22–28 May 2006, pp. 517–520.

Byron, Donna, Alexander Koller, Kristina Striegnitz, Justine Cassell, Robert Dale, Johanna Moore & Jon Oberlander (2009). Report on the First NLG Challenge on Generating Instructions in Virtual Environments (GIVE). In *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009),* Athens, Greece, 30-31 March 2009, pp. 165–173.

Dale, Robert, Sabine Geldof & Jean-Philippe Prost (2005). Using natural language generation in automatic route description. *Journal of Research and Practice in Information Technology*, 37(1):89–106.

Denis, Michel (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Current Psychology of Cognition*, 16:409–458.

Forst, Martin (2007). Filling statistics with linguistics – Property design for the disambiguation of German LFG parses. In *Proceedings of the ACL 2007 Workshop on Deep Linguistic Processing,* Prague, Czech Republic, 28 June 2007, pp. 17–24.

Habel, Christopher (2003). Incremental generation of multimodal route instructions. In Reva Freedman & Charles Callaway (Eds.), *Working Papers of the 2003 AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*, pp. 44–51. Menlo Park, California: AAAI Press.

Kelleher, John D. & Fintan J. Costello (2009). Applying computational models of spatial prepositions to visually situated dialog. *Computational Linguistics*, 35(2):271–306.

Krahmer, Emiel, Sebastiaan van Erk & André Verleg (2003). Graph-based generation of referring expressions. *Computational Linguistics*, 29(1):53–72.

Lemnitzer, Lothar & Claudia Kunze (2002). GermaNet – representation, visualization, application. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation,* Las Palmas, Canary Islands, Spain, 29–31 May 2002, pp. 1485–1491.

Marciniak, Tomacz & Michael Strube (2005). Beyond the pipeline: Discrete optimization in NLP. In *Proceedings of the 9th Conference on Computational Natural Language Learning,* Ann Arbor, Mich., USA, 29–30 June 2005, pp. 136–145.

Maxwell, John T. & Ronald M. Kaplan (1993). The interface between phrasal and functional constraints. *Computational Linguistics*, 19(4):571–590.

Richter, Kai-Florian (2008). *Context-Specific Route Directions – Generation of Cognitively Motivated Wayfinding Instructions*. Amsterdam: IOS Press.

Roth, Michael & Anette Frank (2009). A NLG-based application for walking directions. In *Companion Volume to the Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing,* Singapore, 2-7 August 2009. To appear.

Schmid, Helmut (1997). Probabilistic Part-of-Speech tagging using decision trees. In Daniel Jones & Harold Somers (Eds.), *New Methods in Language Processing*, pp. 154–164. London, U.K.: UCL Press.

Tappe, Heike (2000). Perspektivenwahl in Beschreibungen dynamischer und statischer Wegeskizzen. [Choice of perspective in descriptions of dynamic and static sketchmaps]. In Christopher Habel & Christiane v. Stutterheim (Eds.), *Räumliche Konzepte und sprachliche Strukturen*, pp. 69–97. Tübingen: Niemeyer.

Taylor, Holly & Barbara Tversky (1996). Perspective in spatial descriptions. *Journal of Memory and Language*, 35:371–391.

Viethen, Jette & Robert Dale (2008). The use of spatial relations in referring expression generation. In *Proceedings of the Fifth International Natural Language Generation Conference,* Salt Fork OH, USA, 12–14 June 2008, pp. 59–67.

von Stutterheim, Christiane, Ralf Nüse & Jorge M. Serra (2002). Crosslinguistic differences in the conceptualisation of events. In Hilde Hasselgård, Stig Johansson, Bergljot Behrens & Cathrine Fabricius-Hansen (Eds.), *Information Structure in a Cross-lingustic Perspective*, pp. 179–198. Amsterdam: Rodopi.