Computing EM-based Alignments of Routes and Route Directions as a Basis for Natural Language Generation

Michael Roth and **Anette Frank**
Department of Computational Linguistics
Heidelberg University
`{mroth,frank}@cl.uni-heidelberg.de`

Abstract

Route directions are natural language (NL) statements that specify, for a given navigational task and an automatically computed route representation, a sequence of actions to be followed by the user to reach his or her goal. A corpus-based approach to generate route directions involves (i) the selection of elements along the route that need to be mentioned, and (ii) the induction of a mapping from route elements to linguistic structures that can be used as a basis for NL generation.

This paper presents an Expectation-Maximization (EM) based algorithm that aligns geographical route representations with semantically annotated NL directions, as a basis for the above tasks. We formulate one basic and two extended models, the latter capturing special properties of the route direction task. Although our current data set is small, both extended models achieve better results than the simple model and a random baseline. The best results are achieved by a combination of both extensions, which outperform the random baseline and the simple model by more than an order of magnitude.

1 Introduction

The purpose of route directions is to inform a person, who is typically not familiar with his current environment, of how to get to a designated goal. Generating such directions poses difficulties on various conceptual levels such as planning the route, selecting landmarks (i.e., recognizable buildings or structures) and splitting the task into appropriate single instructions of how to navigate along the route using the selected landmarks as reference points.

Previously developed natural language generation (NLG) systems make use of simple heuristics for the task of content selection for route directions (Dale et al., 2005; Roth and Frank, 2009). In our work, we aim for a corpus-based approach that can be flexibly modeled after natural, human-produced directions for varying subtasks (e.g., indoor vs. outdoor navigation), and that facilitates multilingual extensions. By employing salient landmarks and allowing for variation in NL realization, such a system is expected to generate natural sounding directions that are easier to memorize and easier to follow than directions given by a classical route planner or navigation system.

NLG for route directions crucially differs from other generation tasks such as document summarization (Mani, 2001) in that the selection and ordering of input structures for language generation is heavily situation-dependent, i.e., dependent on the specific properties of a given route to be followed.

In line with a corpus-based NLG approach, we propose to automatically align geographical route representations as produced by a route planner with an annotated corpus of NL directions given by humans for the respective routes. The induced alignments will (i) serve to identify which elements of a route to select for verbalization, and (ii) deliver correspondences between route segments and linguistic input structures that can be used as a basis for statistical NL generation. We investi-

gate a minimally supervised method for inducing such alignments to ensure maximal flexibility for adaptations to different scenarios.

The remainder of this paper is structured as follows: In Section 2 we discuss related work. Section 3 introduces the task, and the representation formats and resources we use. Section 4 introduces a basic Expectation-Maximization model and two extensions for the alignment task. Section 5 outlines the experiments and presents the evaluation results. In Section 6 we conclude and discuss future work.

2 Related Work

Various aspects of route directions have been subject of research in computational linguistics, ranging from instructional dialogues in MapTask (Anderson et al., 1991) to recent work on learning to follow route directions (Vogel and Jurafsky, 2010). However, little work has been done on generating NL directions based on data from Geographical Information Systems (Dale et al., 2005; Roth and Frank, 2009).

NLG systems are typically realized as pipeline architectures (Reiter and Dale, 2000). As a first step, they compute a set of messages that represent the information to be conveyed to a user, given a specific communicative task (*Content Selection*). Selecting appropriate content for a task can be defined heuristically, by manually crafted rules or by learning content selection rules automatically from corpus data. Previous work by Dale et al. (2005) and Roth and Frank (2009) on generating NL directions used hand-crafted heuristics. Duboue and McKeown (2003) were the first to model content selection as a machine learning task, in which selection rules are induced from pairs of human-written text and associated sets of database entries. They induce baseline selection rules from exact matches of NL expressions with database entries; in addition, class-based rules are computed by matching database entry types against NL expressions, using statistical co-occurrence clusters. Barzilay and Lapata (2005) incorporate the interplay between multiple events and entities when learning content selection rules using a special link function.

Recent work by Liang et al. (2009) focuses on modeling grounded language, by aligning real-world representations with NL text that references corresponding world states. They show how a generative model can be used to segment text into utterances and to identify relevant facts with minimal supervision. Both tasks are handled jointly in a unified framework by training a hierarchical semi-Markov model on pairs of text and world states, thereby modeling sequencing effects in the presentation of facts. While their work is not primarily concerned with NLG, the learned correspondences and their probabilities could be applied to induce content selection rules and linguistic mappings in a NLG task. The approach is shown to be effective in scenarios typical for NLG settings (weather forecasts, RoboCup sportscasting, NFL recaps) that differ in the amount of available data, length of textual descriptions, and density of alignments.

In the following, we will adapt ideas from their EM-based approach to align (segments of) route representations and NL route directions in a minimally supervised manner. We will investigate increasingly refined models that are tailored to the nature of our task and underlying representations. In particular, we extend their approach by exploiting semantic markup in the NL direction corpus.

3 Aligning Routes and Directions

In this work we explore the possibility of using an implementation of the EM algorithm (Dempster et al., 1977) to learn correspondences between (segments of) the geographical representation of a route and linguistic instructions of how to follow this route in order to arrive at a designated goal. We are specifically interested in identifying which parts of a route are realized in natural language and which kinds of semantic constructions are used to express them.

As a data source for inducing such correspondences we use a parallel corpus of route representations and corresponding route directions that were collected in a controlled experiment for navigation in an urban street network (cf. Schuldes et al. (2009)). For the alignment task, the routes were compiled to a specification format that has been realized in an internal version of an online route planner. Figure 1 displays the route rep-
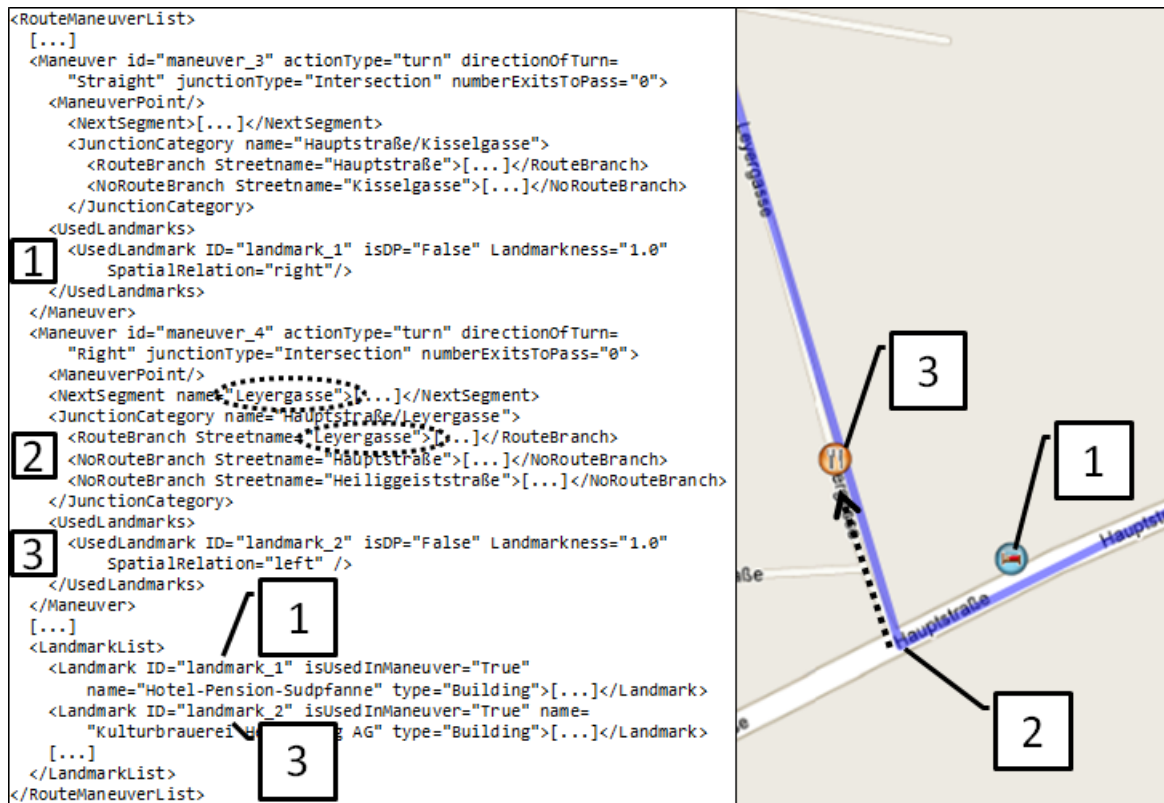
Figure 1: A (partial) route representation of the route segment displayed on the right.

resentation for a small route segment (a junction connecting 'Hauptstraße' and 'Leyergasse'). The corresponding part of a NL route direction is displayed in Figure 2. The route representation and the NL direction share some common concepts: For example, both contain references to a landmark called "Sudpfanne" (marked as [1]) and a street named "Leyergasse" (marked as [2]). Using pairs of route representations and directions, we aim to automatically induce alignments between such correspondences. In the following we describe our data in more detail.

3.1 Route Representation Format

The route representation format we use (illustrated in Figure 1) is an extended version of the *OpenGIS Location Service (OpenLS) Implementation Standards*, a set of XML-based representations specified by the Open Geospatial Consortium[1]. Previous approaches on extending the latter with landmarks in an interopera-

ble way have been presented by Neis and Zipf (2008). The representation format of our data has been developed in close collaboration with researchers from Geoinformatics at Heidelberg University[2] and adopts ideas previously proposed in the *Cognitive OpenLS* specification by Hansen et al. (2006). The resulting specification will be implemented in an extended (internal) version of the online route planner OpenRouteService.org.

Our work revolves around two kinds of elements in this format: so-called *maneuvers*, i.e., elements that describe a decision point including the required action and the following route segment, and *landmarks* that occur along the route. For the alignment task we focus on the following types of attributes that are part of the XML specification, specified here as **Attribute (Element)**:

directionOfTurn (Maneuver) – the direction of movement for the current maneuver, i.e., "left", "right" or "straight"

---

[1]http://www.opengeospatial.org/standards/is

[2]Chair of GIScience, Alexander Zipf, http://www.geog.uni-heidelberg.de/lehrstuehle/gis/
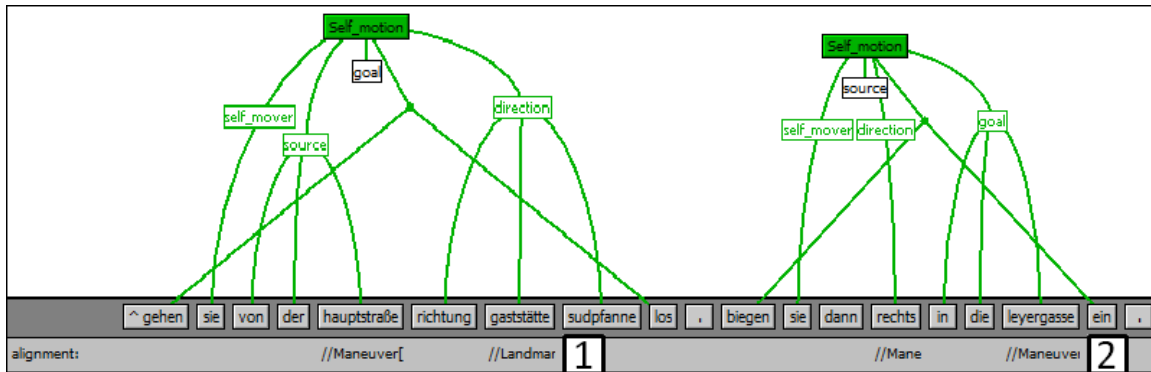
Figure 2: Directions for the route segment displayed in Figure 1 annotated with frame-semantic markup and alignment information. The directions translate to "You start walking from Hauptstraße towards Gaststätte Sudpfanne, then you turn right onto Leyergasse"

junctionType (Maneuver) – the type of junction at the current maneuver, e.g., "intersection", "crossing"

name (JunctionCategory) – the name of the junction at the current maneuver, e.g., "Hauptstraße/Leyergasse"

name (NextSegment) – the name of the street of the next route segment, e.g., "Hauptstraße"

streetName (RouteBranch) – the street name of a branch along which the route continues, e.g., "Leyergasse"

streetName (NoRouteBranch) – the street name of a branch that is not part of the route, e.g., "Kisselgasse"

name (Landmark) – the name of a landmark, e.g., "Hotel Sudpfanne"

spatialRelation (UsedLandmark) – the spatial relation between a landmark and the current maneuver, e.g., "left", "right", "before"

3.2 A Parallel Corpus of Route Directions

The corpus of route directions used in this work is a subset of the data collected by Schuldes et al. (2009) in a desk-based experiment. To elicit NL route directions, subjects were shown a web application that guided them along a route by means of a 2D animation. Subsequently they had to write NL route directions in German for the shown routes. The subjects were allowed to use all information displayed by the web application: named places, buildings, bridges and street names, etc. The resulting directions were POS-tagged with TreeTagger (Schmid, 1997), dependency-parsed with XLE (Maxwell and Kaplan, 1993), and manually revised. Additionally, we annotated frame-semantic markup (Fillmore et al., 2003) and gold standard alignments to the route representation using the SALTO annotation tool (Burchardt et al., 2006).

Frame semantic markup. The texts are annotated with an inventory of 4 frames relevant for directions (SELF_MOTION, PERCEPTION, BE-ING_LOCATED, LOCATIVE_RELATION), with semantic roles (*frame elements*) such as DIREC-TION, GOAL, PATH, LOCATION. Figure 2 illustrates a typical example for the use of the SELF_MOTION frame, once with the elements SOURCE and DIRECTION, and once with the elements DIRECTION and GOAL. Our alignment model uses the frame semantic annotation as structuring information.

Gold standard alignments. For evaluation we constructed gold alignments. We asked two annotators to align text parts with corresponding attributes in the respective route representation[3]. The information about corresponding attributes was added to a single word by manually insert-

---

[3]The alignments have not been double annotated, hence no measure for inter-annotator agreement can be provided.

	#S	#W	#FE	#aligned FE
avg. per direction	8	98	28	14 (50%)
overall	412	5298	1519	750

Table 1: Corpus statistics: number of sentences (S), words (W), frame elements (FE) and alignments.

	#attributes	#aligned attr.
avg. per route	115	14 (12%)
overall	921	

Table 2: Corpus statistics: total number and percentage of relevant attribute alignments.

ing XPATH expressions that unambiguously refer to the aligned attribute in the route representation format. For learning the alignment model, the annotations were spread to all words in the span of the respective frame element.

Corpus statistics. We made use of a corpus of 54 NL directions collected for 8 routes in an urban street network. Tables 1 and 2 give some statistics about the number of words (W) and frame elements (FE) in the parallel corpus. Comparing the total number of relevant attributes (as listed in Section 3.1) and attributes annotated in the gold alignments (aligned attr.) we note that only 12% are actually mentioned in NL directions. Thus it is necessary to select the most salient attributes to avoid the generation of overly redundant text.

4 Alignment Model

For the induction of alignments between (parts of) route structures and semantic representations, we adopt ideas from the models presented in Liang et al. (2009) (cf. Section 2).

We start from a basic *frame alignment model*. It specifies a conditional probability distribution $p(f|a)$ for the alignment to a frame element f of type f_t (e.g., source, goal, direction) in the frame-semantic annotation layer given an attribute a of type a_t (e.g., streetName, directionOfTurn) in the route representation format. Note that this model does not take into account the actual value a_v of the attribute a nor the words that are annotated as part of f. We assume that the frame annotation represents a reliable segmentation for this alignment. This allows us to omit modeling segmentation explicitly.

As extensions to the basic frame alignment model, we specify two further models that capture properties that are specific to the task of direction alignment. As route directions are typically presented in a linear order with respect to the route, we incorporate an additional *distance model* λ in our alignment. We further account for *word choice* within a frame element as an additional factor. The word choice model $p(w|a)$ will exploit attribute type and value information in the route representations that are reflected in word choice in the linguistic instructions. Both extensions are inspired by and share similarities with models that have been successfully applied in work on text alignment for the task of machine translation (Vogel et al., 1996; Tiedemann, 2003).

Our full model is a distribution over frame elements f and words w that factorizes the three above mentioned parts under the assumption of independence between each component and each attribute:

$$p(f, w|a) = p(f|a)\lambda(\text{dist}(f, a))\,p(w|a) \quad (1)$$

The individual models are described in more detail in the following subsections.

4.1 Frame Alignment Model

This basic frame alignment model specifies the probabilities $p(f|a)$ for aligning an attribute a of type a_t (i.e., one of the types listed in Section 3.1) to a frame element f labeled as type f_t. This alignment model is initialized as a uniform distribution over f and trained using a straight-forward implementation of the EM algorithm, following the well-known IBM Model 1 for alignment in machine translation (Brown et al., 1993). The expectation step (E-step) computes expected counts given occurrences of f_t and a_t under the assumption that all alignments are independent 1:1 correspondences:

$$\text{count}(f_t, a_t) = \frac{\sum_{\{\langle f', a'\rangle | f'_t = f_t \wedge a'_t = a_t\}} p(f'|a')}{\sum_{\{\langle f', y\rangle | f'_t = f_t\}} p(f'|y)} \quad (2)$$

The probabilities are re-estimated to maximize the overall alignment probability by normalizing

the estimated counts (M-step):

$$p(f|a) = \frac{\text{count}(f_t, a_t)}{\sum_x \text{count}(x_t, a_t)} \qquad (3)$$

4.2 Distance Model

We hypothesize that the order of route directions tends to be consistent with the order of maneuvers encoded by the route representation. We include this information in our alignment model by defining a distance measure $\text{dist}(f, a)$ between the relative position of a frame element f in the text and the relative position of an attribute a in the route representation. The probabilities are specified in form of a distance distribution $\lambda(i)$ over normalized distances $i \in [0 : 1]$ and learned during EM training. The weights are initialized as a uniform distribution and re-estimated in each M-step by normalizing the estimated counts:

$$\lambda(i) = \frac{\sum_{\{\langle x,y \rangle \,|\, \text{dist}(x,y)=i\}} \text{count}(x, y)}{\sum_{\{\langle x,y \rangle\}} \text{count}(x, y)} \qquad (4)$$

4.3 Word Choice Model

We define a word choice model for word usage within a frame element. This additional factor is necessary to distinguish between various occurrences of the same type of frame element with different surface realizations. For example, assuming that the frame alignment model correctly aligns directionOfTurn attributes to a frame element of type DIRECTION, the word choice model will provide an additional weight for the alignment between the value of an attribute (e.g., "left") and the corresponding words within the frame element (e.g., "links"). Similarly to the word choice model within fields in (Liang et al., 2009), our model specifies a distribution over words given the attribute a. Depending on whether the attribute is typed for strings or categorial values, two different distributions are used.

String Attributes. For string attributes, we determine a weighting factor based on the longest common subsequence ratio (LCSR). The reason for using this measure is that we want to allow for spelling variants and the use of synonymous common nouns in the description of landmarks and street names (e.g., "Main St." vs. "Main Street",

"Texas Steakhouse" vs. "Texas Restaurant"). The weighting factor $p_{\text{str}}(w|a)$ for an alignment pair $\langle f, a \rangle$ is a constant in the E-step and is calculated as the LCSR of the considered attribute value a_v and the content words $w = \text{cw}(f)$ in an annotated frame element f divided by the sum over the LCSR values of all alignment candidates for a:

$$p_{\text{str}}(w|a) = \frac{\text{LCSR}(a_v, w)}{\sum_x \text{LCSR}(a_v, \text{cw}(x))} \qquad (5)$$

Categorial Attributes. We define categorial attributes as attributes that can only take a finite and prescribed set of values. For these we do not expect to find matching strings in NL directions as the attribute values are defined independently of the language in use (e.g., values for directionOfTurn are "left", "right" and "straight". However, the directions in our data set are in German, thus containing the lexemes "links", "rechts" und "geradeaus" instead). As the set of values $\{a_v \in \mathbb{D}_{a_t}\}$ for a categorial attribute type a_t is finite, we can define and train probability distributions over words for each of them during EM training. The models are initialized as uniform distributions and are used as a weighting factor in the E-Step. We re-calculate the parameters of a distribution $p_{cat}(w|a)$ in each EM iteration by normalizing the estimated counts during M-step:

$$p_{\text{cat}}(w|a) = \frac{\text{count}(a_v, w)}{\sum_x \text{count}(a_v, x)} \qquad (6)$$

5 Experiments and Results

5.1 Setting

We test the performance of different combinations of these EM-based models on our data, starting from a simple baseline model (**EM**), combined with the distance (**EM+dst**) and word choice models (**EM+wrd**) and finally the full model (**Full**). We perform additional experiments to examine the impact of different corpus sizes and an alignment threshold (**+thld**).

EM is a baseline model that consists of a simple EM implementation for aligning attributes and frame elements (equation (3)).

EM+dst consists of the simple EM model and the additional distance factor (equation (4)).

Model	P (+thld)	R (+thld)	F_1 (+thld)
Random	2.7 (2.7)	3.9 (3.9)	3.2 (3.2)
EM	2.0 (3.6)	2.9 (3.7)	2.34 (3.6)
EM+dst	7.3 (11.6)	10.8 (11.7)	8.7 (11.6)
EM+wrd	26.8 (36.3)	39.5 (35.5)	32.0 (35.9)
Full	28.9 (38.9)	42.5 (37.9)	34.4 (38.4)

Table 3: Precision (P), Recall (R) and F_1 measure results with and without threshold (+thld) on the alignment task (all numbers in percentages).

EM+wrd consists of the simple EM model with the word choice model (equations (5) and (6), respectively).

Full is the full alignment model including distance and word choice as described in Section 4 (cf. equation (1)).

We use the data set described in Section 3. The predictions made by the different models are evaluated against the gold standard alignments (cf. Tables 1 and 2). We run a total number of 30 iterations[4] of EM training on the complete data set to learn the parameters of the probability distributions. From the set of all possible 1-to-1 alignments, we select the most probable alignments according to the model in a way that no attribute and no frame element is aligned twice.

5.2 Results

We measure precision as the number of predicted alignments also annotated in the gold standard divided by the total number of alignments generated by our model. Recall is measured as the number of correctly predicted alignments divided by the total number of alignment annotations. As baselines we consider a random baseline (obtained from the average results measured over 1,000 random alignment runs) and the simple EM model.

The results in Table 3 show that the simple EM model performs below the random baseline. The individual extended models achieve significant improvement over the simple model and the random baseline. While the distance model has a smaller impact, the influence of the word choice

---

[4]This number was determined by experiments as a general heuristics.

# directions	Precison	Recall	F_1
1	28.94%	42.31%	34.38%
2	29.04%	41.90%	34.31%
3	29.01%	42.18%	34.38%
4	28.75%	41.81%	34.07%
5	29.36%	42.69%	34.79%
6	30.18%	43.91%	35.77%

Table 4: Average results when using only a specific number of directions for each route with the model Full (-thld).

model is considerable. Applying the full model yields further performance gains. We note that for all models recall is higher compared to precision.

One of the reasons for this phenomenon may be that the EM-based models align as many attributes as possible to frame elements in the route directions. In our gold standard, however, only around 12% of all relevant attributes correspond to frame elements in the route directions (cf. Section 3.2). We estimate this quota from a part of the corpus and use it as an alignment threshold, i.e., for evaluation we select the best alignments proposed by the models, until we reach the threshold. With this we achieve a F_1 measure of 38.40% in a 6-fold cross validation test. This represents an improvement of 3.97 points and considerably boosts precision, yielding overall balanced precision (38.90%) and recall (37.92%).

A general problem of the current setup is the small amount of available data. With a total of 54 route directions, the data consists of 6 to 8 directions for each route. We compute a learning curve by using only exactly 1 to 6 directions per route to examine whether performance improves with increasing data size. The results are computed as an average over multiple runs with different data partitions (see Table 4). The results indicate small but consistent improvements with increasing data sizes, however, the differences are minimal. Thus we are not able to conclude at this point whether performance increases are possible with the addition of more data.

5.3 Error Analysis

In an error analysis on the results of the full model, we found that 363 out of 784 (46%) misalign-

ments are related to attributes not aligned in our gold standard. This is due to the fact that not all relevant attributes are realized in natural language directions. By addressing this problem in the model Full+threshold, we are able to reduce these errors, as evidenced by a gain of almost 10 points in precision and 4 points in F_1 measure.

We further observe that the word choice model does not correctly reflect the distribution of categorial attributes in the parallel corpus. In the data, we observe that humans often aggregate multiple occurrences of the same attribute value into one single utterance. An example of such a phenomenon can be seen with the attribute type 'directionOfTurn': Even though "straight" is the most common value for this attribute, it is only realized in directions in 33 (5%) cases (compared to 65% and 47% for "left" and "right" respectively). While our EM implementation maximizes the likelihood for all alignment probabilities based on expected counts, many pairs are not – or not frequently – found in the corpus. This results in the model often choosing incorrect alignments for categorial attributes and makes up for 23% of the misaligned attributes in total.

We found that further 5% of the attributes are misaligned with frame elements containing pronouns that actually refer to a different attribute. As our word choice model does not account for the use of anaphora, none of the affected frame elements are aligned correctly. Given the genre of our corpus, integrating simple heuristics to resolve anaphora (e.g., binding to the closest preceding mention) could solve this problem for the majority of the cases.

6 Conclusion

We presented a weakly supervised method for aligning route representations and natural language directions on the basis of parallel corpora using EM-based learning. Our models adopt ideas from Liang et al. (2009) with special adaptations to the current application scenario. As a major difference to their work, we make use of frame-semantic annotations on the NL side as a basis for segmentation.

While we can show that the extended models significantly outperform a simple EM-based

model, the overall results are still moderate. We cannot draw a direct comparison to the results presented in Liang et al. (2009) due to the different scenarios and data sets. However, the corpus they used for the NFL recaps scenario is the closest to ours in terms of available data size and percentage of aligned records (in our case attributes). For this kind of corpus, they achieve an F_1 score of 39.9% with the model that is closest to ours (Model 2'). Their model achieves higher performance for scenarios with more available data and a higher percentage of alignments. Thus we expect that our model benefits from additional data sets, which we plan to gather in web-based settings.

Still, we do not expect to achieve near to perfect alignments due to speaker variation, a factor we also observe in the current data. As our ultimate goal is to generate NL instructions from given route representations, we can nevertheless make use of imperfectly aligned data for the compilation of high-confidence rules to compute semantic input structures for NLG. Following previous work by Barzilay and Lee (2002), we can also exploit the fact that our data consists of multiple directions for each route to identify alternative realization patterns for the same route segments. In addition, (semi-)supervised models could be used to assess the gain we may achieve in comparison to the minimally supervised setting.

However, we still see potential for improving our current models by integrating refinements based on the observations outlined above: Missing alignment targets on the linguistic side – especially due to anaphora, elliptical or aggregating constructions – constitute the main error source. We aim to capture these phenomena within the linguistic markup in order to provide hidden alignment targets. Also, our current model only considers frame elements as alignment targets. This can be extended to include their verbal predicates.

References

Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry Thompson, and Regina Weinert. 1991. The HCRC Map Task corpus. *Language and Speech*, 34(4):351–366.

Barzilay, Regina and Mirella Lapata. 2005. Collective content selection for concept-to-text-generation. In *Proceedings of the Human Language Technology Conference and the 2005 Conference on Empirical Methods in Natural Language Processing,* Vancouver, B.C., Canada, 6–8 October 2005, pages 331–338.

Barzilay, Regina and Lillian Lee. 2002. Bootstrapping lexical choice via multiple-sequence alignment. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing,* Philadelphia, Penn., 6–7 July 2002, pages 164–171.

Brown, Peter F., Vincent J. Della Pietra, Stephan A. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19:263–311.

Burchardt, Aljoscha, Katrin Erk, Anette Frank, Andrea Kowalski, and 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, pages 517–520.

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

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistics Society, Series B (Methodological)*, 39(1):1–38.

Duboue, Pablo A. and Kathleen R. McKeown. 2003. Statistical acquisition of content selection rules. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing,* Sapporo, Japan, 11–12 July 2003, pages 121–128.

Fillmore, Charles J., Christopher R. Johnson, and Miriam R.L. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography*, 16(3):235–250.

Hansen, Stefan, Kai-Florian Richter, and Alexander Klippel. 2006. Landmarks in OpenLS: A data structure for cognitive ergonomic route directions. In *Proceedings of the 4th International Conference on Geographic Information Science,* Münster, Germany, 20-23 September 2006.

Liang, Percy, Michael Jordan, and Dan Klein. 2009. Learning semantic correspondences with less supervision. In *Proceedings of ACL-IJCNLP 2009*, pages 91–99, August.

Mani, Inderjeet. 2001. *Automatic Summarization*. John Benjamins, Amsterdam, Philadelphia.

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

Neis, Pascal and Alexander Zipf. 2008. Extending the OGC OpenLS route service to 3D for an interoperable realisation of 3D focus maps with landmarks. *Journal of Location Based Services*, 2(2):153–174.

Reiter, Ehud and Robert Dale. 2000. *Building Natural Language Generation Systems*. Cambridge, U.K.: Cambridge University Press.

Roth, Michael and Anette Frank. 2009. A NLG-based Application for Walking Directions. In *Companion Volume to the Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing,* Singapore, 2–7 August 2009, pages 37–40.

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

Schuldes, Stephanie, Michael Roth, Anette Frank, and Michael Strube. 2009. Creating an annotated corpus for generating walking directions. In *Proceedings of the ACL-IJCNLP 2009 Workshop on Language Generation and Summarisation,* Singapore, 6 August 2009, pages 72–76.

Tiedemann, Jörg. 2003. Combining Clues for Word Alignment. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 339–346, Budapest, Hungary.

Vogel, Adam and Dan Jurafsky. 2010. Learning to Follow Navigational Directions. In *Proceedings of ACL-2010*, Uppsala, Sweden.

Vogel, Stephan, Hermann Ney, and Christoph Tillmann. 1996. HMM-based Word Alignment in Statistical Translation. In *Proceedings of the 16h International Conference on Computational Linguistics (COLING)*, pages 836–841, Copenhagen, Denmark.