

Analysing ParCor and its Translations by State-of-the-art SMT Systems

Liane Guillou

School of Informatics
University of Edinburgh
Scotland, United Kingdom
L.K.Guillou@sms.ed.ac.uk

Bonnie Webber

School of Informatics
University of Edinburgh
Scotland, United Kingdom
bonnie@inf.ed.ac.uk

Abstract

Previous work on pronouns in SMT has focussed on third-person pronouns, treating them all as anaphoric. Little attention has been paid to other uses or other types of pronouns. Believing that further progress requires careful analysis of pronouns as a whole, we have analysed a parallel corpus of annotated English-German texts to highlight some of the problems that hinder progress. We combine this with an assessment of the ability of two state-of-the-art systems to translate different pronoun types.

1 Introduction

Previous work on the translation of pronouns in Statistical Machine Translation (SMT) has focussed on the specific problem of translating *anaphoric* pronouns – i.e., ones that co-refer with an *antecedent* entity previously mentioned in the discourse (Le Nagard and Koehn, 2010; Hardmeier and Federico, 2010; Guillou, 2012; Novák et al., 2013; Hardmeier, 2014; Weiner, 2014). This is because languages differ in how an anaphoric pronoun relates to its antecedent, and the relationship does not fit naturally into the SMT pipeline. Some pronoun forms also have non-anaphoric uses, and there are other types of pronouns. Languages also differ as to what types of pronouns are used for what purposes.

To investigate similarities and differences in pronoun usage across languages, we conducted an analysis of the ParCor corpus¹ of pronoun annotations over a set of parallel English-German texts. The corpus contains a collection of texts from two different genres: 8 EU Bookshop² publications (written text) and 11 TED³ Talks (tran-

scribed planned speech). In the ParCor annotations, each pronoun is marked as being one of eight *types*: Anaphoric/cataphoric, event reference, extra-textual reference, pleonastic, addressee reference, speaker reference, generic reference, or other function⁴. Additional features are recorded for some pronoun types, for example anaphoric/cataphoric pronouns are linked to their antecedents. Full details of the annotation scheme are provided in Guillou et al. (2014).

Through analysing similarities and differences in pronoun use in these parallel texts, we hope to better understand the problems of translating different types of pronouns. This knowledge may in turn be used to build discourse-aware SMT systems in the future. In addition, through analysing translations produced by state-of-the-art systems, we hope to understand how well current systems translate a range of pronoun types. This information may be used to identify the pronoun types where future efforts would be best directed.

The advantage of using the ParCor corpus is that it allows us to conduct part of the analyses automatically once we have word-aligned the parallel texts. The annotations also allow for the separation of ambiguous pronouns such as “it” which may serve as an anaphoric, event or pleonastic pronoun⁵. This allows for a more granular analysis than has been provided in other similar studies.

2 Previous Work

There has been previous work both on comparing pronoun usage in English and German (in the genre of business letters using comparable rather than parallel texts (Becher, 2011) and for the multi-genre GECCo corpus (Kunz and Lapshinova-Koltunski, 2015)) and on pronoun translation accuracy by SMT systems (Hardmeier and Federico, 2010; Novák et al., 2013;

¹<http://opus.lingfil.uu.se/ParCor/>

²EU Bookshop: <https://bookshop.europa.eu>

³TED: WIT³ corpus: <https://wit3.fbk.eu/>

⁴Pronoun does not belong to any of the other categories

⁵Each pronoun type has different translation requirements

Pronoun Type	TED Talks				EU Bookshop			
	English		German		English		German	
<i>Anaphoric</i>	886	(27.71)	1,228	(40.52)	2,767	(20.32)	3,036	(22.72)
Anaphoric (pronominal adverb)	N/A		N/A		70	(0.51)	84	(0.63)
Cataphoric	5	(0.16)	16	(0.53)	67	(0.49)	19	(0.14)
Event	264	(8.26)	331	(10.92)	239	(1.76)	255	(1.91)
Event (pronominal adverb)	N/A		N/A		0	(0.00)	78	(0.58)
Extra-textual reference	52	(1.63)	26	(0.86)	N/A		N/A	
<i>Pleonastic (non-referential)</i>	61	(1.91)	224	(7.39)	191	(1.40)	391	(2.93)
Addressee reference	499	(15.61)	525	(17.32)	112	(0.82)	76	(0.57)
Speaker reference	1,386	(43.35)	1,467	(48.41)	548	(4.02)	580	(4.34)
Generic	N/A		N/A		9	(0.07)	58	(0.43)
Pronoun (other)	N/A		N/A		135	(0.99)	126	(0.94)
Pronoun (unsure)	N/A		N/A		14	(0.10)	0	(0.00)
Total	3,153	(98.62)	3,817	(125.95)	4,152	(30.49)	4,703	(35.20)

Table 1: Pronoun **type** counts for English (source) and German (translation) texts in ParCor. Counts per 1000 tokens are provided in parentheses. *N/A* indicates that the type is not marked for one of the corpora

Weiner, 2014), these being relatively small scale. The main focus, however, has been on building models to improve pronoun translation in SMT through targeting different stages of the translation process. These include pre-annotation of the source-language data (Le Nagard and Koehn, 2010; Guillou, 2012), decoder features (Hardmeier and Federico, 2010; Novák et al., 2013; Hardmeier, 2014; Weiner, 2014) and post-editing / re-ranking (Weiner, 2014). Despite these efforts, little progress has been made.

In the most comprehensive study to date, Hardmeier (2014) concludes that current models for pronoun translation are insufficient and that “...*future approaches to pronoun translation in SMT will require extensive corpus analysis to study how pronouns of a given source language are rendered in a given target language*”. This paper reports on such a corpus analysis.

3 Analysis of Manual Translation

Identifying and understanding systematic differences in pronoun use between a pair of languages may help inform the design of SMT systems. With this in mind, we compared original English texts and their human-authored German translations in the ParCor corpus, for both genres, at the corpus, document and sentence levels.

3.1 Corpus-level

Corpus-level comparison reveals the first differences between pronoun use in the two languages. (See Table 1. Some counts differ from those in (Guillou et al., 2014) due to minor changes

prior to corpus release and the automatic addition of first person pronouns and German “man”.) Specifically, the German translations contain more anaphoric and pleonastic pronouns than the original English texts. (A *pleonastic* pronoun does not refer to an antecedent, e.g. “**It** is raining” / “**Es** regnet”.) Paired t-tests show that this difference is significant for pleonastic pronouns in both the TED corpus, $t(10)=-5.08$, $p < .01$, and the EU Bookshop corpus, $t(10)=-3.68$, $p < .01$. The difference in anaphoric pronoun use is significant for the TED corpus, $t(7)=-3.52$, $p < .01$, but not the EU Bookshop corpus, $t(7)=-1.09$, ($p=0.31$).

3.2 Document-level

Again, at the document-level we observe that the German translations typically contain more anaphoric and pleonastic pronouns than the original English texts. (See Table 2 for the pronoun counts of a randomly selected document, 767.)

Pronoun Type	English	German
<i>Anaphoric</i>	121 (22.53)	189 (39.58)
Cataphoric	0 (0.00)	2 (0.42)
Event	49 (9.12)	59 (12.36)
Extra-textual ref.	5 (0.93)	6 (1.26)
<i>Pleonastic</i>	8 (1.68)	54 (11.31)
Addressee reference	102 (18.99)	91 (19.06)
Speaker reference	156 (29.04)	163 (34.14)
Pronoun (unsure)	3 (0.56)	0 (0.00)
Total	444 (82.67)	564 (118.12)

Table 2: Pronoun **type** counts for TED Talk 767. Counts per 1000 tokens provided in parentheses

Similar trends were observed for the other documents in the corpus which suggests that this is

not simply a consequence of stylistic differences over authors or speakers. A presentation of the full analysis would, however, require a longer paper.

Documents in ParCor were originally produced in English and then translated into German. To ascertain whether similar patterns of pronoun use can be observed for the opposite translation direction, we annotated two German TEDx talks and their English translations, again using the guidelines described in Guillou et al. (2014).

We observed similar patterns, with more pleonastic pronouns used in German than in English (19 vs. 11 pleonastic pronouns in one document, and 15 vs. 2 in the other). For anaphoric pronouns, one document has 119 in the German original and 140 in the English translation, with near equal numbers (54 vs. 51) in the other document. With only two documents it is not possible to confirm whether German systematically makes use of more anaphoric and pleonastic pronouns, but cf. Becher (2011) who points to several patterns, in particular the insertion of explicit possessive pronouns in English-to-German translation and pronominal adverbs in the opposite direction.

3.3 Sentence-level

Pronoun counts at the corpus and document levels are simply raw counts. They do not tell us anything about cases in which a pronoun is used in the original text and dropped from the translation (deletions), or is absent from the original text but present in the translation (insertions). To discover this, we need to drill down to the sentence-level.

We start with the sentence-aligned parallel texts provided as part of the ParCor release. In order to identify the German translation of each pronoun in the original English text, we compute word alignments using Giza++ (<https://code.google.com/p/giza-pp/>) with *grow-diag-final-and* symmetrisation. To ensure robust alignments, we concatenated the ParCor texts and additional data – specifically, the IWSLT 2013 shared task training data (for TED and TEDx) and Europarl data (for EU Bookshop). We consider an English and German pronoun to be equivalent if the following conditions hold: (a) a word alignment exists between them, and (b) they share the same pronoun type label in the ParCor annotations.

To evaluate the word-alignment quality we examined a random sample of 100 parallel sentences from the TED corpus. The sentences contain 213

English and 241 German pronouns. We define a bad alignment as one where a pronoun is aligned to something that is not the corresponding pronoun in the other language, or should be unaligned but is not. We find that 6.57% of English and 9.12% of German pronouns are part of a bad alignment.

Taking TED talk 767 as an example and using the combination of pronoun type and alignments to identify a source-target pronoun match, we observe many mismatches. Table 3 shows that 412 pronouns are unique to either the English original or the German translation, with only 298 matching English-German pronoun pairs. The largest absolute difference lies in the number of anaphoric pronouns in the target for which there is no comparable pronoun in the source (*anaphoric* insertions), followed by *pleonastic* insertions.

Pronoun Type	English (deletion)	German (insertion)
<i>Anaphoric</i>	49	117
Cataphoric	0	2
Event	26	36
Extra-textual ref.	4	5
<i>Pleonastic</i>	3	49
Addressee reference	31	20
Speaker reference	30	37
Pronoun (unsure)	3	0
Total	146	266

Table 3: Sentence-level pronoun **type + alignment** mismatches for TED Talk 767

There is no single reason for *anaphoric* deletions: Anaphoric pronouns may be omitted from the German output for stylistic reasons, as a result of paraphrasing or possibly to conform with language-specific constraints. With respect to *anaphoric* insertions, intra-sententially, many correspond to relativizers in English. That is, while in English a relative clause is introduced with a *that-*, *wh-* or *null-relativizer*, an anaphoric pronoun serves as a relativizer in German.⁶ For example, “that” in “The house **that** Jack built” is a relativizer and the corresponding “das” in “Das Haus, **das** Jack gebaut hat” is a relative pronoun. Manual analysis of the German translation for TED Talk 767 identified 42 cases where an anaphoric pronoun was inserted as a relative pronoun corresponding to a relativizer in English. While this does not explain all of the *anaphoric* insertions, it is frequent enough to deserve further attention.

⁶The ParCor corpus has not marked instances of *that* when used as a relativizer in English.

Several fixed expressions in English appear to trigger *pleonastic* insertions in German. A commonly observed pair is “There +*be*”/“Es gibt”. These *existential there* constructions are not annotated in ParCor, but their presence accounts for some (not all) of the insertions of pleonastic pronouns in German. As the fixed expressions are short and occur frequently, phrase-based systems could be expected to provide accurate translations.

3.4 Discussion

We have observed differences in pronoun use in both genres of the ParCor corpus. Since SMT systems are trained on parallel data similar to that in ParCor, it is important to be aware that content words such as nouns and verbs are more likely to be faithfully translated as there are fewer ways to convey the same meaning. On the other hand, there is more variation in the translation of function words such as pronouns — for example in active to passive conversions (and vice versa). Where there is a lot of variation the SMT system may not be able to learn accurate mappings.

To this is added the problem of ambiguous pronouns such as “it”, for which the anaphoric and pleonastic forms both translate as “es” in German. These frequent alignments in the training data may also bias the likelihood that “it” is incorrectly translated as “es” (neuter), even if a feminine or masculine pronoun is required in German.

4 Assessing Automated Translation

Analyses of the output of state-of-the-art SMT systems provide an indication of how well current systems are able to translate pronominal coreference — what they are good and bad at. We follow our analysis of manual translation and examine English-to-German translation for anaphoric pronouns (“it” and “its”) and relativizers.

For our state-of-the-art systems, we selected two systems from the IWSLT 2014 shared task in machine translation (Birch et al., 2014). The first is a phrase-based system that incorporates factored models for words, part-of-speech tags and Brown clusters. The second is a syntax-based, string-to-tree, system. Both systems were trained using a combination of TED data and corpora provided for the WMT shared task. Here, TED talks are considered to be in-domain, with the EU Bookshop texts considered out-of-domain.

We are not interested in making direct compar-

isons between the two systems, as their different training makes such comparisons unfair. However, similarities in the translation accuracy of two systems can show that our findings are not specific to a single system or type of system.

For manual translation, we can assume that a pronoun is accurately translated, inserted or dropped, as part of a close translation of the original sentence or an acceptable paraphrase. As such, it is reasonable to use automated analysis based on the ParCor annotations and alignments between the texts. With automated translations, however, there is no guarantee that a source pronoun is translated correctly by the system. We therefore need to rely more heavily on manual analysis.

However, manual analysis can be aided by some automated pre-processing steps, to help select pronouns for further study. Using the source text and its translation together with word alignments output by the SMT systems, we can investigate which pronouns may be more difficult to translate than others – i.e. we can produce frequency distributions of the translations produced for each source pronoun surface-form (split by pronoun type).

4.1 Identifying Pronouns for Analysis

Examining the translation frequency distributions for the two state-of-the-art systems, we can observe the following. First, “it” can be translated into German, depending on the context, as either masculine singular (sg.), feminine sg. or neuter sg., or plural. As plural pronouns are not gendered, “they” has fewer translations. The possessive pronoun “its” has additional possible translation options due its multiple dependencies. That is, possessive pronouns in German must agree in number/gender with both the possessor and the object that is possessed. Different base forms are used depending on whether the possessor is feminine/plural (“ihr”) or masculine/neuter (“sein”). Other anaphoric pronouns such as “he” and “she” have far fewer translation options and are therefore less interesting. Based on the possible translation options, we selected (anaphoric) “it” and “its”.

Our analysis of manual translation (Section 3.3) showed that relativizers in English often corresponded to a relative pronoun inserted in the German translation. We wish to see how well SMT systems handle the translation of relativizers. We selected that-relativizers (explicit in English text) and null-relativizers (implicit). We exclude wh-

relativizers, also explicit, but with many forms (what, who, etc.), to reduce the annotation effort.

4.2 Pronoun Selection Task

Our manual analysis of pronoun translation is framed as a pronoun selection task. In this setting a human annotator is asked to identify which pronoun(s) could validly replace a placeholder masking a pronoun at a specific point in the SMT output. By masking the pronoun, we remove the risk that the annotator is biased by the pronoun present in the SMT output. The annotator’s selections may then be compared with the pronouns produced by the system in order to assess translation accuracy.

We used the tool described by Hardmeier (2014) for the pronoun selection task. The interface presents the annotator with the source sentence and its translation plus up to five previous sentences of history, as well as a number of pronoun options. The source pronoun in the final sentence of each example block is highlighted and its translation is replaced with a placeholder.

To determine how many sentences of history to present to the annotator (to help them identify the antecedent of an anaphoric pronoun), we used the manual annotations in ParCor. We calculated both the mean number of sentences between a pronoun and its antecedent, and two standard deviations from the mean (accounting for 95% of pronouns). (Intra-sentential pronouns have a distance of zero.) For the TED corpus the mean distance between pronoun and antecedent is 1.33 sentences, and two standard deviations from the mean is 4.95 sentences. For the EU Bookshop (whose sentences are longer), the distances between pronoun and antecedent are typically shorter, with a mean distance of 0.67 sentences and two standard deviations from the mean at 3.57 sentences. We nevertheless allow for up to five previous sentences of history for each example, regardless of genre.

4.3 Pronoun Selection Task: Guidelines

The following guidelines were adapted from those used by Hardmeier (2014) in order to cater for the requirements of English-German translation:

1) Select the pronoun that will create the most fluent translation, while preserving the meaning of the English sentence as much as possible. The latter means assigning correct number/gender to the pronoun that replaces the placeholder: Its case may be left “unknown”.

- If the SMT output is sufficiently fluent to be able to determine the case of the pronoun, select the appropriate check-box.
- Use the plural options if the antecedent is translated as a plural, or in any other scenarios in which a plural might seem appropriate.
- If different, equally grammatical options are available, select all appropriate check-boxes.

2) Alternatively select “Other” if the sentence should be completed with a pronoun not included in the list, “Bad translation” if a grammatical and faithful translation cannot be created without making major changes to the surrounding text, or “Discussion required” if you are unsure what to do.

3) Ignore minor disfluencies (e. g., incorrect verb agreement or obviously missing words).

4) Always try to select the pronoun that best agrees with the antecedent in the SMT output, even if the antecedent is translated incorrectly, and even if this forces you to violate the pronoun’s agreement with immediately surrounding words such as verbs, adjectives etc.

5) If the translation does not contain a placeholder, but a pronoun corresponding to the one marked in the English text should be inserted somewhere, indicate which pronoun should be inserted.

6) If the SMT output does not contain a placeholder, but already includes the correct pronoun, annotate the example as if a placeholder were present. This will mean selecting the same pronoun that is included in the SMT output.

4.4 Anaphoric “it”

The anaphoric pronoun “it” can co-refer either intra-sententially (i.e., to an antecedent in the same sentence) or inter-sententially (i.e., to an antecedent in a different sentence). While co-reference imposes number–gender constraints on a pronoun and its antecedent, intra-sentential coreference imposes additional constraints.

We randomly selected 50 inter- and 50 intra-sentential tokens of “it” labelled anaphoric in the ParCor annotations. Tokens were selected from the TED Talks, as sentences there are typically shorter than those in the EU Bookshop and hence, potentially easier to work with. Additional guidelines are provided for “it”:

- Select “Pronominal adverb” if the most fluent translation would come from using a Ger-

man pronominal adverb⁷. (Selection of the pronominal adverb is not required.)

- If a demonstrative pronoun (e.g. “diese” or “jene”) is possible, select whether it is **more** or **less** likely than the personal pronoun(s).
- Genitive options are not available as these are used for possessives.

The annotator is presented with a table of options for number/gender and case combinations. The number/gender options are masculine, feminine, neuter and plural. The case options are: “case unknown”, and three German cases: nominative, accusative and dative. See Figure 1.

Select the correct pronoun:

	Masculine	Feminine	Neuter	Plural
Case unknown	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Nominative	<input type="checkbox"/> er	<input type="checkbox"/> sie	<input type="checkbox"/> es	<input type="checkbox"/> sie
Accusative	<input type="checkbox"/> ihn	<input type="checkbox"/> sie	<input type="checkbox"/> es	<input type="checkbox"/> sie
Dative	<input type="checkbox"/> ihm	<input type="checkbox"/> ihr	<input type="checkbox"/> ihm	<input type="checkbox"/> ihnen

Demonstrative pronoun (e.g. diese/jene) possible (personal pronoun **more** likely)
 Demonstrative pronoun (e.g. diese/jene) possible (personal pronoun **less** likely)

Figure 1: Annotator options for “it”

Although the ParCor annotations contain antecedent links for anaphoric pronouns, we did not display these to the annotator for any of the tasks.

4.5 Anaphoric possessive “its”

In German, *dependent* possessive pronouns (i.e. those that precede a noun) must agree not only with the number/gender of its antecedent (possessor) but also with the number/gender of its object (i.e. the noun that follows the pronoun). For example in: “**Der** Staat und **seine** Einwohner” (“The state and its inhabitants”) the antecedent “Staat” (“state”) is masculine (sg.) and so a “sein” form is required for the possessive pronoun. The ending “e” in “seine” is needed because the noun following the possessive pronoun is plural (“Einwohner/inhabitants”).

We randomly selected 50 instances of “its” marked as anaphoric in ParCor. As “its” is uncommon in the TED corpus, all 50 instances came from the EU Bookshop corpus. Additional guidelines are provided for “its”:

- Select the relevant combination of number/gender of possessor and object. Select the

⁷Pronominal adverbs also exist in English (e.g. therefore, wherein, hereafter) but are used more frequently in German

case of the pronoun if the quality of the SMT output permits this.

- Select “Pronoun not required” if the translation does not require a pronoun.

The annotator is presented with a table of options capturing the number/gender of the possessor vs. the number/gender of the object. To reduce the number of options, a separate set of check-boxes is provided for case options, including “case unknown”, nominative, accusative, dative and genitive.

4.6 Relativizers

English relativizers may be explicit (*that-* and *wh-* relativizers), or implicit (*null-relativizers*). Both may be translated as relative pronouns in German.

We randomly selected 50 instances of relativizers from the TED corpus; 25 *that-* and 25 *null-relativizers*. The selection was semi-automatic, based on identifying relative clauses in the output of the Berkeley Parser (Petrov et al., 2006) and manually selecting those that contained a *that-* or *null-relativizer*.

As *null-relativizers* are implicit, there are no tokens in the English text to highlight. To keep this task in line with the others, we manually insert symbols for the nulls, i.e. the “∅” in “The house ∅ Jack built”, and (manually) align them to the corresponding token in the SMT output. (Unalignable tokens are left untranslated.) Instead of a pronoun in the English text, the annotator is presented with an instance of “that” or a symbol representing the *null-relativizer*. Placeholders are included in the translation as normal.

The options table captures pronoun number/gender and case. It is similar to the table for “it”, but with relative pronoun forms and options for “case unknown” and all four German cases.

5 Results

The results of the three pronoun selection tasks are presented in Table 4. We automatically compared the translations produced by the systems with the selections made by the annotator. If the system-generated pronoun matches one of the annotator’s selections, there is a “pronoun match”. If it doesn’t match any of the annotator’s selections or the system did not generate a pronoun there is a “pronoun mismatch”. Matches are recorded in terms of number/gender and case if the annotator supplied it, or number/gender only, if not.

Result	“it”				“its”		Relativizers			
	Inter		Intra		PB	Syn	That		Null	
	PB	Syn	PB	Syn			PB	Syn	PB	Syn
Pronoun match (number/gender + case)	20	8	14	15	15	9	14	12	13	12
Pronoun match (number/gender only)	0	1	1	0	8	10	0	0	2	0
Pronoun mismatch	14	28	27	26	24	28	2	3	1	1
Pronoun not translated (mismatch)	1	0	0	0	0	0	4	3	5	6
Pronominal adverb match	5	8	2	2	N/A	N/A	N/A	N/A	N/A	N/A
Pronominal adverb mismatch	2	0	0	1	N/A	N/A	N/A	N/A	N/A	N/A
Other	2	0	2	1	0	0	2	3	3	3
Bad translation	4	1	1	2	1	1	3	4	1	2
Pronoun not required	0	1	0	0	2	2	0	0	0	1
Anaphoric but could not find antecedent	0	1	0	0	0	0	0	0	0	0
Unsure: may not be anaphoric	2	2	3	3	0	0	0	0	0	0
Sub-total	50	50	50	50	50	50	25	25	25	25
Total	200				100		100			

Table 4: Pronoun selection task results for anaphoric “it”, anaphoric possessive “its” and relativizers. PB=Phrase-based system, Syn=Syntax-based system, Inter=pronoun and antecedent are not in the same sentence. Intra=pronoun and antecedent in same sentence. Pronominal adverb is an option for “it” only

For most examples the annotator was able to determine the case of the pronoun as well as its number/gender. Recall that the annotator was specifically instructed to only select the case of the pronoun if the SMT output was sufficiently fluent so as to make this possible. It would therefore appear that our initial assumption that it might be difficult to identify syntactic role was not entirely correct.

“Pronominal adverb match” is used when the SMT output contains a pronominal adverb and the annotator had indicated that one would be appropriate. As the annotator was not asked to specify the pronominal adverb, we make no further comparison. “Pronominal adverb mismatch” is the opposite; the annotator indicated that a pronominal adverb should be used but the system did not output one. “Other”, “Bad translation” and “Pronoun not required”⁸ are used for those pronouns marked as such in the pronoun selection task.

Some instances of “it” were initially left for discussion. These were later assigned one of two new categories: “Anaphoric but could not find antecedent” where the antecedent could not be identified due to insufficient history or “Unsure: may not be anaphoric” where the annotator believed that the pronoun may not in fact be anaphoric, despite being labelled as such in the ParCor corpus.

Instead of comparing the systems, we use the results from both to assess how well state-of-the-art systems perform at pronoun translation. We find that both systems typically produce more incorrect translations than correct ones.

⁸Although “pronoun not required” was not initially provided for the “it” task, we added it later when the need arose.

Both systems regularly translate “it” as “es”: 79/100 cases for the phrase-based and 78/100 for the syntax-based system. This reflects biases in the training data, where the use of “it” and “es” as both anaphoric and pleonastic pronouns leads to their frequent alignment. A similar bias is observed for relativizers, with both that- and null-relativizers commonly translated as “die”. For example, both systems translate “that” as “die” in 13 of the 21 instances in which a translation is provided, though not the same 13 of 21 instances.

It is often acceptable to translate “it” using either a personal or demonstrative pronoun: 49/100 cases for the phrase-based and 59/100 cases for the syntax-based system. However, neither system generated demonstrative pronouns, perhaps due to the bias toward translating “it” as “es”.

For “its” the systems often select an incorrect base form for the pronoun: i.e. “ihr” when “sein” should be used, and vice versa. The phrase- and syntax-based systems selected the incorrect base form for 17/50 and 15/50 instances respectively.

Both systems are able to insert relative pronouns when a null-relativizer is encountered in the English source text, with a similar accuracy to the translation of that-relativizers. One might have expected that translating an explicit source token would be easier (and more accurate) than inserting a token in the SMT output which has no explicit representation in the source.

6 Discussion: Anaphoric “it”

When annotating the English side of ParCor, deciding whether a pronoun was anaphoric, event-

related or pleonastic was one of the major causes of annotator disagreement. It is therefore not surprising that problems might arise in identifying the pronoun’s antecedent for the pronoun selection task. This ambiguity did not arise for the “its” or relativizers tasks. With “its”, events are rarely (if ever) possessors and so rarely serve as antecedents. With relativizers, the relative pronoun and its antecedent (in German) are likely to be very close together, and certainly intra-sentential.

The syntax-based system is much better at translating intra-sentential pronouns than inter-sentential ones. Although this system contained no such enhancements, one might expect that pronoun-aware syntax-based systems could be designed to leverage the fact that intra-sentential pronouns are syntactically governed, and produce better translations. One possible option would be to combine two systems: a phrase-based system to translate inter-sentential pronouns, and an enhanced syntax-based system to translate intra-sentential pronouns.

7 Discussion: Relative pronouns

When the antecedent is not a noun, i.e. “something” (“etwas”), “anything” (“alles”/“jedes” etc.) or “nothing” (“nichts”), “was” should be used:

- (1) Now , when I use the term miracle , I don ’t mean something **that** ’s impossible.
- (2) Nun , wenn ich den Begriff Wunder verwenden , ich meine nicht etwas , **XXX** ist unmöglich .

As “was” is not provided as an option in the pronoun selection task, the annotator marked example 2 (and others like it) as “other”. SMT systems must decide whether to use a relative pronoun that conveys the number/gender of the antecedent (i.e. der/die/das) or “was/wer/wo” (if the antecedent cannot be determined / there is no antecedent). As this decision depends on the antecedent, relative pronouns may therefore be treated as a more localised sub-set of anaphoric pronouns.

The translation of relativizers may require a preposition preceding the relative pronoun:

- (3) That ’s the planet \emptyset we live on .
- (4) Das ist die Welt , **XXX** wir leben .

The correct translation of example 3, which contains a null-relativizer (indicated by \emptyset), would be “Das ist die Welt, **in der** wir leben”. However,

in the SMT output the preposition “in” is missing, and so the annotator was required to select the correct pronoun as if the preposition had been present.

In German, the choice of preposition and case of the pronoun are determined by the verb of the clause. As these choices are connected, SMT systems could also consider the translation of prepositions when translating relative pronouns.

8 Conclusion

The analysis of manual translation revealed that pronouns are frequently dropped and inserted by human translators and that German translations contain many more pleonastic and anaphoric pronouns than the original English texts. Both of these differences can result in SMT systems learning poor translation mappings.

The analysis of state-of-the-art translation revealed that biases in the training data and incorrect selections of the base form pronoun (i.e. “ihr” vs. “sein” for “its”) are both problems which SMT systems must overcome. For relative pronouns selecting the correct preposition is also important as it influences the case of the pronoun.

9 Future Work

Possible directions for future work include further analyses of manual and automated translation and applying the knowledge that is gained to build pronoun-aware SMT systems. Initial efforts could focus on syntax-based SMT — leveraging information within target-side syntax trees constructed by the decoder, to encourage pronoun-antecedent agreement for intra-sentential anaphoric pronouns (i.e. “it/its” and relative pronouns).

Pronoun-aware SMT systems could also address translation of the ambiguous second-person pronouns “you” and “your”. In English, they have both deictic and generic use, while in German, different forms are used (“Sie/du” vs. “man”).

Acknowledgements

Thanks to Christian Hardmeier for providing the pronoun-selection tool, to Susanne Tauber for completing the annotations and to Felix Suessenbach and Sam Gibbon for assisting with the significance testing. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement n° 645452 (QT21).

References

- Viktor Becher. 2011. *Explicitation and implicitation in translation. A corpus-based study of English-German and German-English translations of business texts*. Ph.D. thesis, Department of Applied Linguistics (Institut für Sprachlehrforschung), University of Hamburg.
- Alexandra Birch, Matthias Huck, Nadir Durrani, Nikolay Bogoychev, and Philipp Koehn. 2014. Edinburgh SLT and MT System Description for the IWSLT 2014 Evaluation. In *International Workshop on Spoken Language Translation*, pages 49–56, Lake Tahoe, CA, USA.
- Liane Guillou, Christian Hardmeier, Aaron Smith, Jörg Tiedemann, and Bonnie Webber. 2014. ParCor 1.0: A Parallel Pronoun-Coreference Corpus to Support Statistical MT. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Liane Guillou. 2012. Improving Pronoun Translation for Statistical Machine Translation. In *Proceedings of the Student Research Workshop, EACL 2012*, pages 1–10.
- Christian Hardmeier and Marcello Federico. 2010. Modelling Pronominal Anaphora in Statistical Machine Translation. In *Proceedings of the 7th International Workshop on Spoken Language Translation*.
- Christian Hardmeier. 2014. *Discourse in Statistical Machine Translation*. Ph.D. thesis, Department of Linguistics and Philology, Uppsala University.
- Kerstin Kunz and Ekaterina Lapshinova-Koltunski. 2015. Cross-linguistic analysis of discourse variation across registers. *Nordic Journal of English Studies*, 14(1):258–288.
- Ronan Le Nagard and Philipp Koehn. 2010. Aiding Pronoun Translation with Co-Reference Resolution. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 252–261.
- Michal Novák, Anna Nedoluzhko, and Zdeněk Žabokrtský. 2013. Translation of “It” in a Deep Syntax Framework. In *Proceedings of the 1st Workshop on Discourse in Machine Translation, ACL 2013*, pages 51–59.
- Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein. 2006. Learning Accurate, Compact, and Interpretable Tree Annotation. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics, ACL-44*, pages 433–440, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Jochen Weiner. 2014. Pronominal Anaphora in Machine Translation. Master’s thesis, Karlsruhe Institute of Technology, Karlsruhe, Germany.