Effect of Information Packaging on Perceived Translation Quality

Susan Howlett Mark Dras

June 8, 2010

Abstract

This report describes the purpose, construction, and results of the survey entitled *Effect of Information Packaging on Perceived Translation Quality*. We tested the hypothesis that native German speakers prefer a translation into German where the same element is prominent as in the English sentence over a translation where a different element is prominent, and found statistically significant support for our hypothesis. We note some potential difficulties with our assumptions and recommend that our findings be corroborated with a different language pair.

1 Introduction

Information packaging (IP) is the order and manner in which information is presented in a sentence, a choice between alternatives made available by the grammar of a language. The purpose of this study was to investigate native German speakers' preferences regarding translations where different IP choices have been made when translating out of English.

1.1 Information Packaging in English

English predominantly uses *intonation* to signal IP. However, other methods are also available, through syntactic constructions such as *clefts*. The following example from Huddleston and Pullum (2002, p. 1414) illustrates three different kinds of cleft in comparison with the unmarked word order.

Unmarked I bought a red wool sweater.

it-**cleft** It was a red wool sweater that I bought.

wh-**cleft** What I bought was a red wool sweater.

Reverse *wh*-**cleft** A red wool sweater was what I bought.

In this survey, we restrict our attention to *it*-clefts and *wh*-clefts only.

Huddleston and Pullum (2002) describe clefts as dividing the content of the unmarked clause into two parts, one *foregrounded* and the other *backgrounded*. We label these two parts X and Y respectively. With these labels we can describe the form of *it*-clefts as *It is/was X that Y* and *wh*-clefts as *What Y is/was X*. We assume that the function of a cleft is to give prominence to the element X.

1.2 Information Packaging in German

Sentences in German may be divided into three sections—the Vorfeld (VF), Mittelfeld (MF) and Nachfeld (NF). The form and function of these sections are described in Section H of Lederer (1969), where the terms are translated as prefield, sentence field and postfield, respectively.

The MF is delimited by the finite element of the predicate on the left and all remaining parts of the predicate (which may be null) on the right. There is a canonical order for constituents within the MF, but other orders are possible, reflecting the significance of the information (Lederer, 1969, H700–4). The VF comes before the MF and contains at most one constituent; the NF likewise consists of at most one constituent after the MF. Filippova and Strube (2007) cite research that the sentence-initial position is cognitively prominent, and claim that the VF is therefore the ideal location for either the topic of the sentence or a "frame-setting" element. In other words, the VF can have an IP role, as does the order of constituents in the MF. We assume that placing an element in the VF lends prominence to that element.

Finally, German also makes use of IP constructions similar to English, including clefts. The structure of the German cleft parallels that of the English. We assume that the German cleft gives prominence to the same element of the sentence as the parallel English cleft.

2 Hypothesis

The aim of the survey was to test the hypothesis that, given an English sentence, native German speakers prefer a translation into German where the same element is prominent as in the English sentence over a translation where a different element is prominent, using our assumptions about prominence outlined in Section 1.

3 Survey

The survey consisted of 40 questions presented in random order. Each question consisted of an English sentence, with between one and four sentences of prior context in English[1], plus two alternative German translations, in random order. Participants were asked to choose which of the two alternative translations they preferred, or to select 'no preference'. There was also an optional field for the participants to make additional comments.

Potential questions were selected from the German–English Europarl training data from the 2009 Workshop on Statistical Machine Translation (WMT'09),[2] before tokenisation, through two regular expressions:

- `It (is|was)( [^ ]+){1,3} (that|who|which)`
- `What( [^ ]+){1,6} (is|was).`[3]

From the English sentences that matched these regular expressions, 20 *it*-clefts and 20 *wh*-clefts were chosen by hand. Half of each kind had the cleft removed to create an unmarked word order with neutral emphasis.

[1] The variability in context length was chiefly due to difference in sentence length, but shorter contexts were also used where a new topic had only just been introduced.

[2] http://www.statmt.org/wmt09/translation-task.html

[3] The choice of the numbers 3 and 6 was arbitrary.

Subject	Type	n_+	n_0	n_-
1	Neutral	10	3	7
	it-cleft	6	1	3
	wh-cleft	4	3	3
2	Neutral	12	3	5
	it-cleft	5	0	5
	wh-cleft	7	0	3
3	Neutral	7	8	5
	it-cleft	5	3	2
	wh-cleft	5	3	2
Total		61	24	35

Table 1: Breakdown of responses by question type.

We manually created two alternative German translations for each English sentence based on the reference translation in the corpus.[4] Alternatives were constructed by reordering components of the sentence so that a different element was in the VF or in the *X* position of a cleft. In questions where the English sentence had been de-clefted, one alternative placed emphasis on an element of the sentence and the other had neutral emphasis, as judged by a native German speaker who did not participate in the experiment. In questions where the cleft remained in the English sentence, the two sentences placed emphasis on two different elements. If our hypothesis is correct, we would expect participants to choose the option where the emphasis is the same between the two sentences.

The full list of questions and details of how each question was derived from corpus data are given in the accompanying files, `questions.txt` and `question-history.txt`. The file `responses.txt` contains the responses from each participant. Each row corresponds to a participant and each column to a question. A cell contains `+1` if the participant chose the translation that supported the hypothesis for that question, `-1` if he or she chose the other translation, and `0` if he or she chose no preference.

4 Results

Three native German speakers completed our survey. All were university students or staff and fluent in English.

Raw numbers for each participant are given in Table 1, subdivided according to the question type: *Neutral* for the 20 questions where the English sentence was de-clefted, *it-cleft* for the 10 questions where the English sentence contained an *it*-cleft, and *wh-cleft* for the 10 *wh*-cleft questions. The n_+ column indicates the number of times each participant chose the response that agreed with our hypothesis. Similarly, n_- indicates the number of times they chose the other alternative, while n_0 indicates the number of times they selected 'no preference'.

From the breakdown by question type, there does not appear to be a difference in how participants responded for the different kinds of clefts.

[4] In 33 of the 40 questions, the reference translation was used verbatim as one alternative.

To test for significance, we use an extension of the sign test for replicated measurements, outlined in Woodbury, Manton, and Woodbury (1977). We use the normal approximation for their test, given as

$$z = \frac{(n_+ + n_0/2) - n/2 - 0.5}{\sqrt{((1/12)\sum_{i=1}^{k} r_i(r_i+2)) + n_0/4}}$$

where z is a z-score, and r_i is the number of replicates (judges) for the ith of k independent questions. In our data, $k = 40$, and $r_i = 3$ for all i.

Using this test on the totals given in Table 1, we obtain a z-score of 1.67, giving a p-value of 0.0475, significant at $\alpha = 0.05$. While this is close to the significance boundary, we note that it was obtained with only three judges.

5 Discussion

The statistical significance of our results provides some support for our hypothesis that native speakers prefer a translation where the same element is prominent as in the original sentence. The fact that the level of significance is marginal is largely a reflection of the small number of judges, but also suggests that perhaps the view of IP as a means for making an element prominent is incomplete and/or that the assumptions about the placement of prominence are inaccurate.

During the construction of the survey, it became evident that the placement of prominence in *wh*-clefts is not clear cut. The question was raised of whether some part of *Y* received prominence instead of, or perhaps in addition to, *X*. This was particularly true in cases like the following, where there may be emphasis on *criminal* rather than *the behaviour ... for illegal purposes*.

(1) What is criminal is the behaviour of some 3 to 5% of Internet users who misuse the Net for illegal purposes.

A more refined analysis of *wh*-clefts is clearly required, possibly incorporating multiple prominent locations. Our analysis of *it*-clefts, however, appears reasonable.

With this imbalance in validity of assumptions between *it*-clefts and *wh*-clefts, it is perhaps surprising that there is no discernible difference between *it*-cleft and *wh*-cleft questions in the survey results. Most likely, this is due to the limited amount of data, both number of participants and number of questions. However, other aspects of the survey construction may have contributed.

Although we avoided asking participants questions based on the uncertain prominence in a *wh*-cleft (for example, in the sentence given above, the two alternatives emphasised *the behaviour ... for illegal purposes* and *some 3 to 5% ... for illegal purposes*, not *criminal*), we did not actively or consistently control for other possible sources of prominence in the sentences, for example words like *precisely* and the German equivalent *genau*, or influences of context.

We may also have inadvertently confused the issue by relying on a German native speaker's judgements about emphasis in addition to our assumed prominent locations. These alternative sources and interpretations of emphasis may have added noise to the data, which could help to mask any effect present.

Finally, it is important to note that because German and English are closely-related languages, it is frequently the case that a simple literal or close-to-literal translation produces

an acceptable result. It would be necessary to repeat this experiment with a less closely-related language pair, where this effect is smaller, in order to verify our findings.

6 Conclusion

We constructed a survey to test the hypothesis that native speakers prefer translations that maintain the same prominence as the original sentence over a translation where a different element is made prominent. We examined the translation of *it*-clefts and *wh*-clefts from English into German. Our findings supported our hypothesis at a significance level of $\alpha = 0.05$.

However, the construction of the survey highlighted some difficulties with our assumptions regarding prominence in *wh*-clefts. This, along with some of our decisions in constructing the survey and the choice of language pair, may have contributed some noise to the findings. We recommend that the IP assumptions underlying our survey be re-examined and our findings corroborated with a larger sample size in a different language pair.

References

Filippova, Katja and Michael Strube. 2007. The German Vorfeld and local coherence. *Journal of Logic, Language and Information*, 16(4):465–485.

Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge University Press.

Lederer, Herbert. 1969. *Reference grammar of the German language: Based on Grammatik der deutschen Sprache by Dora Schulz and Heinz Griesbach*. Scribner.

Woodbury, Max A., Kenneth G. Manton, and Lowell A. Woodbury. 1977. An extension of the sign test for replicated measurements. *Biometrics*, 33(3):453–461, September.