IBM Research Report

A Translation Confidence Index for English-German MT

Arendse Bernth , Claudia Gdaniec

IBM Research Division
Thomas J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598

IBM

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

A Translation Confidence Index for English-German MT

Arendse Bernth and Claudia Gdaniec [1]

Abstract. The quality of Machine Translation output can vary a great deal from sentence to sentence, even for the same MT system. A facility for automatically rating the confidence in the MT system's translation of each segment would vastly improve the usability of the MT system. Obviously, such a confidence index must depend in part on the differences between the source and target languages. In this paper we outline the nature of a translation confidence index (the TCI), which is tunable to different language-pair MT systems, and we then focus on the interesting contrastive linguistic issues in tuning the TCI to the case of English-German.

1 INTRODUCTION

The quality of output produced by a Machine Translation (MT) system can vary a great deal. As argued in [2], an indication of the likely quality of the translation output greatly improves the usability of an MT system because such a measure makes it possible to filter out or highlight possible bad translations. [1] described the overall design of a translation confidence index, hereinafter called the *TCI*, and [3] described the contribution of source analysis, which plays a crucial role in many MT systems, including ours.

The TCI is fully implemented and has been successfully integrated with the IBM English-German MT system [4, 8] for which it has been tuned to a precision of around 75% for the TCI's quality judgments.

The TCI is based on the idea of measuring the complexity of the translation process. As the complexity increases, the confidence in the translation decreases. We start out with a perfect score of 100, which gets penalized as various situations arise. This is described in more detail in [1, 3].

The potential problems that may arise during the translation process have different impacts depending on the language pair in question. This is reflected in a language-pair-specific profile where the various potential problems are assigned the penalties that are used for that particular language pair. Proper calibration of the language-pair-specific profile is a matter of great importance. In this paper we shall give some technical details about those aspects of the TCI that relate to translation from English to German. This is of general interest for the NLP community because:

1. It shows how to make a real-life instantiation of a general and useful scheme of meta-level reasoning in a way that is illustrative and useful for others for their own applications of this scheme.
2. We address issues of general interest for MT between English and German, which is one of the most desired language pairs.

This IBM MT system is a transfer-based system, and that is reflected in the design of the TCI. Such a system falls naturally into three stages: source analysis, transfer, and target generation. During these phases, hooks into the MT system at crucial points call routines that look for the various types of problems, assign penalties, and calculate intermediate values for the TCI. Our focus in this paper is on transfer, even though we will also consider source analysis briefly, as it relates to English-German.

2 SOURCE ANALYSIS

Source analysis plays a prominent role in any transfer-based MT system, and the impact of various types of problems on the TCI is described in detail in [3]. Here we shall only describe a few high-impact problems that may be encountered during source analysis and only those of particular concern for translating from English to German.

2.1 Incomplete parses

The source analysis is produced by the English Slot Grammar (ESG) [5, 6, 7]. The greatest problem that may occur during source analysis is obviously if the sentence cannot be given a complete parse. Such sentences are given a pieced-together parse. Whereas it is true that incomplete parses will affect translation into other languages besides German, it has a rather high impact for English to German, due to the types of differences between those languages. Where English only has agreement requirements for subject-verb configurations, German also has agreement requirements for e.g. noun phrases and pronouns. In order to get the agreement right, it is important to have a reliable modifier structure (as expressed by a good parse). In addition to agreement, there is the issue of word order, which can be very different in English and German. Structural transfer (see Section 3) works by tree transformations on the structure produced initially by the parser. So a well-formed parse is also crucial in this context. We assign a penalty of 50 for a segment that cannot be given a complete parse.

2.2 Ing-verbs

Another problem is the occurrence of ing-verbs in the source text, because these can be used in so many ways in English, few of which have a correspondence in the German present participle. The uses of ing-forms in English and the corresponding ways of expressing them in German include the following:[2]

1. **English:** *Following* that step, hosts receive a confirmation note.
   German: *Nach* diesem Schritt erhalten Gastgeber eine Bestätigungsnotiz.
   (IBM translation: Nach diesem Schritt erhalten Gastgeber eine Bestätigungsnotiz. TCI score: 83.80)

---

[1] IBM T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, USA, email: arendse@us.ibm.com and cgdaniec@us.ibm.com

[2] One treatment of ing-forms in English-German MT is described in [10].

2. **English:** By *relying* on an electronic medium, we can eliminate unnecessary paper.
   German: Indem *wir uns* auf ein elektronisches Medium *verlassen*, können wir unnötiges Papier vermeiden.
   (IBM translation: Indem wir uns auf ein elektronisches Medium verlassen, können wir unnötiges Papier entfernen. TCI score: 87.20)
3. **English:** Developers can integrate advanced speech recognition capabilities *using* established graphical tools, *reducing* the need for special *coding*.
   German: Entwickler können hochentwickelte Spracherkennungsanwendungen *mit Hilfe* herkömmlicher graphischer Werkzeuge integrieren *und damit* den Bedarf an spezieller *Verschlüsselung* reduzieren.
   (IBM translation: Entwickler können hochentwickelte Spracherkennungsfähigkeiten mit Hilfe herkömmlicher graphischer Werkzeuge integrieren und reduzieren den Bedarf an spezieller Verschlüsselung. TCI score: 57.26)
4. **English:** A collaboration of seven European nations created the physics laboratory called CERN, CERN being the French acronym for European Center for Nuclear Research.
   German: Die Zusammenarbeit von sieben europäischen Nationen schuf das Physiklabor namens CERN, wobei CERN das französische Akronym für Europäisches Zentrum für Kernforschung ist.
   (IBM translation: Eine Kollaboration sieben europäischer Nationen schuf das Physiklabor, das CERN, CERN heißt, das das französische Akronym für europäisches Zentrum für Kernforschung ist. TCI score: 11.31)

Ing-verbs are penalized 5 per occurrence.

2.3 Passives

While simple passives do not pose any problems for source analysis itself they do pose problems for transfer because of the various ways passives can be translated into German.

1. **English:** He *was injured* on the job.
   German: Er *verletzte sich* auf der Arbeit.
   (IBM translation: Er verletzte sich auf dem Job. TCI score: 83.58)
2. **English:** He *was injured* by the car.
   German: Er *wurde* vom Auto *verletzt*.
   (IBM translation: Er wurde vom Auto verletzt. TCI score: 83.83)
3. **English:** The company has *been talked* about.
   German: Über die Firma wurde gesprochen/Man sprach über die Firma.
   (IBM translation: Über die Firma ist geredet worden. TCI score: 82.05)
4. **English:** This software *was designed* to help end-users.
   German: Diese Software *war dazu gedacht*, Endanwendern zu helfen.
   (IBM translation: Diese Software war dafür entworfen, Endbenutzern zu helfen. TCI score: 82.73)
5. **English:** This software *was designed* in a hurry to help end-users.
   German: Diese Software *wurde* in Eile *entworfen*, *um* Endanwendern zu helfen.
   (IBM translation: Diese Software wurde in aller Eile dafür entworfen, Endbenutzern zu helfen. TCI score: 80.25)
6. **English:** Gore *was expected* to win the primaries.
   German: Es wurde erwartet, daß Gore die Vorwahlen gewinnt/Man erwartete, daß Gore die Vorwahlen gewinnt.
   (IBM translation: Es wurde erwartet, daß Gore die Vorwahlen gewinnt. TCI score: 81.68)
7. **English:** Although the fossilized dinosaurs are thought to have been capable of running swiftly, they were unable to fly.
   German: Obwohl man annimmt, daß die versteinerten Dinosaurier schnell laufen konnten, konnten sie nicht fliegen.
   (IBM translation: Obwohl es gedacht ist, daß die versteinerten Dinosaurier fähig sind, schnell zu laufen, waren sie außerstande zu fliegen. TCI score: 72.61)

Passives are penalized 15 per occurrence.

3 TRANSFER

Transfer is divided up into two distinct phases: Lexical transfer and structural transfer. Structural transfer is the phase where one encounters most of the problems that are specific to the translation between English and German, and we will concentrate on this phase of transfer in this paper.

Lexical transfer converts the source parse tree into a transfer tree by replacing the source nodes with target nodes. To the target nodes are attached target word senses, appropriate target features, and in some cases triggers for associated lexical transformations. The target tree structure is largely isomorphic at this stage to that of the source sentence.

Structural transfer applies *transformations* to the target tree to change the order of the nodes, and to add or delete nodes as necessary. This phase represents on a very concrete level the differences between the source and target languages; the transformations are very specific to the language pair in question.

For certain transformations, the conditions for applying the transformation may be complex, or the interaction with other transformations may be complicated. We shall describe here difficulties in the proper interpretation of English coordination, "for-to" constructions, and one case of ing-form transformations.

Some transformations have to process a tree where nodes that have to be made adjacent in German are far apart. If the connection between the distant nodes is not made, the output is unacceptable. However, if one of these transformations succeeds, we have more confidence in the quality of the translation, and a *reward* is given (this is done by applying a negative penalty). We describe one such case. If they succeed only partially, the output improves minimally but points are nevertheless deducted because the system knows that either the parse is incorrect or a transformation was missed.

3.1 Coordination

Due to lack of inflection and agreement in English, modification in coordination of noun phrases is ambiguous: [3]
   English: They planned the update dates and schedules.
   German: Sie planten die Aktualisierungsdaten und -pläne.
   (TCI score: 88.44)

Analysis of the scope of modifiers before coordinated nouns is relatively difficult because it involves complex semantic processing. But German has to make it transparent in different ways: insertion of determiners to agree with gender, number and case of the heads, insertion of hyphens after modifiers or before heads, correct combining forms for modifiers, and inflection of adjectives. Therefore, if the analysis is incorrect or too vague, the result is an ungrammatical

---

[3] Below, all German translations shown are those produced by the IBM English-German MT system.

German phrase. Application of the *Modifiers-in-Coordination* transformations therefore carries with it a penalty of at least 2 points with compounding of penalties if there are more than two coordinated nodes.

English: They planned the update schedules, dates, and personnel.

German: Sie planten die Aktualisierungspläne, die Daten und das Personal.

(TCI score: 79.03).

(Instead of: "Sie planten die Aktualisierungspläne und -daten und das Aktualisierungpersonal.")

Multiple coordination receives a heavier penalty because increasing complexity increases the potential for errors. There are several transformations involved in the treatment of coordination, and they get penalized individually; the total range of penalties for coordination varies from 2 to 11.

3.2   Ing-clause as adverbial

Ing-constructions in English can function as adverbial modifiers to the main clause. They do not contain an overt subject, nor do they have to contain a conjunction which would indicate the type of adverbial, i.e. causal, temporal, purpose, etc.[4] Humans can infer both from the context; a machine cannot. There may be some clues that can be used to infer the kind of adverbial. Unless a reliable clue in the sentence is encountered, the transformation is penalized with a default 10 points, in addition to the 5 points for an ing-form from source analysis.

In the first example, the perfective is an indication that the relationship is temporal, and the transformation can therefore insert a specific temporal conjunction. The subject is copied from the subject of the main clause and replaced there with an anaphoric pronoun. The system is then relatively confident that the output will be correct, and no further penalties are applied.

English: Having finished her homework, Anne went out of the house.

German: Nachdem Anne ihre Hausaufgaben beendet hatte, ging sie aus dem Haus.

(TCI score: 81.57)

In the following example, the type of adverbial can be guessed on the basis of the verb. Certain verbs seem to imply a causal relationship. When they are encountered, the transformation will not be penalized.

English: Feeling tired, Anne went to bed.

German: Weil Anne sich müde fühlte, ging sie ins Bett.

(TCI score: 91.46)

Some other verbs, on the other hand, indicate with relative certainty that the relationship is not distinctly causal. This certainty is also rewarded by not penalizing the sentence additionally to the source penalties:

English: Hearing strange noises downstairs, Tim got nervous.

German: Als Tim seltsame Geräusche unten hörte, wurde er nervös.

(TCI score: 86.00)

When no clues are detected in the sentence, a default transformation applies that adds either "als", "wenn", or "indem", depending on the tense and mood of the main clause. This decision is appropriate in many instances, but it can be wrong, as can be seen in the following example. Since there is little confidence in the quality of the output, 10 points will be deducted.

English: Hoping to raise themselves in society, they are now selling antique furniture.

German: Indem sie hoffen, in der Gesellschaft aufzusteigen, verkaufen sie jetzt antike Möbel.

(TCI score: 65.04)

(Instead of: " In der Hoffnung, in der Gesellschaft aufzusteigen, verkaufen sie jetzt antike Möbel.")

The following sentence receives a very low score, which predicts correctly the bad translation. The score reflects a deduction – among other penalties – of 50 points for an incomplete parse, as well as another 20 points for an unanalyzed node in the tree.

English: It being Christmas, all offices were closed.

German: Es, Weihnachten zu sein, wurden alle Büros geschlossen.

(TCI score: 15.32)

(Instead of: "Da es Weihnachten war, waren alle Büros geschlossen.")

In the following sentence, the ing-form has an instrumental meaning ("by shifting"). This cannot be determined by transformations; therefore the resulting translation is unacceptable, and the score is very low (penalties are applied for other complexities, as well).

English: The new Web site is designed to better involve viewers through a new look and structure, shifting from a vertical, scrolling screen to an enhanced, easy-to-use 640X480 design.

German: Die neue Web-Site ist dafür entworfen, Zuschauer besser durch einen neuen Blick und Struktur zu betreffen und sich von einem senkrechten, blätterndem Bildschirm zu einem verbesserten, benutzerfreundlichen 640X480 Entwurf zu verschieben.

(TCI score: 2.75)

3.3   For-to constructions

The English "for-to" construction is very ambiguous structurally as will be seen from the following examples. It therefore gets a penalty of 5 points when it is detected in the source analysis. As a default, it gets an additional 10 points in transformations, unless one of a particular series of transformations is applied that is deemed correct.

The first example is a straight-forward purpose clause (paraphrase: "In order for the software to run, you need to install it first"). The purpose clause precedes the main clause, and the applied transformation assumes a confidence of correct analysis and transfer. So the sentence does not receive an additional penalty.[5]

English: For the software to run, you need to install it first.

German: Damit die Software läuft, müssen Sie es zuerst installieren.

(TCI score: 89.64)

The second example illustrates how the same construction has a different syntactic function, that of an extraposed non-finite subject clause (paraphrase: "It is possible that my son drives you to the airport"). Here, no such confidence exists, therefore the sentence will be further penalized.

English: It is possible for my son to drive you to the airport.

German: Es ist möglich, daß mein Sohn Sie zum Flughafen fährt.

(TCI score: 81.08)

The following illustrates a completely different syntactic structure. The "for-to" consists of the beneficiary argument of the main predicate while the non-finite clause is the subjectless extraposed subject

---

[4] [9] gives an extensive description of the many, varied uses and meanings of non-finite adverbial clauses.

[5] We assume that a sentence such as *For her to care for her old parents is not easy* is odd.

(paraphrase: "Caring for her old parents is not easy for her"). Although the transformation here decides that the *for*-phrase refers, in fact, to the beneficiary, it is not absolutely confident. Therefore, it still adds the default penalty for "for-to" constructions.

English: It is not easy for her to care for her old parents.
German: Es ist nicht leicht für sie, für ihre alten Eltern zu sorgen.
(TCI score: 77.76)

The last example, finally, shows the "for-to" construction as a modifier of the predicate noun (paraphrase: "It is the rule that children be quiet").

English: There is a rule/It is the rule for children to be quiet.
German: Es gibt eine Regel/Es ist die Regel, daß Kinder ruhig sind.
(TCI score: 86.68/85.47)

Several transformations convert one single syntactic structure in the source into multiple different German patterns. Since this construction is so ambiguous and the transformations have to disambiguate it, the danger that an incorrect resolution occurs is great.

3.4 Unanalyzed things

A special class of transformations are applied at the very end after all other transformations have been tried. This is a set of transformations that try to catch unanalyzed phenomena and avoid the worst output. Their function is to attempt to correct the worst output by setting some default. One example is number and person assignment to verbs. If a verb goes to the target generation component of the system without any number or person marking, it cannot be generated successfully. Therefore, a default number and person is assigned here to guarantee an acceptable verb form. These transformations assume either a wrong source analysis, incorrect or missed transformations, or possibly ungrammatical input.

The following sentence gets heavy penalties for multiple coordination and ing-forms. The second ing-form is transformed incorrectly. The last ing is not processed at all by transformations, except by one of the final clean-up series, which changes the present participle into an infinitive. This is a default transformation which is based on the assumption that a German infinitive – albeit far from correct – is still less unacceptable than a participle. A penalty of 20 points is, nevertheless, deducted because it is clear that the analysis or the transfer process has failed somehow.

English: In his spare time, Andrew enjoys travelling, sculpting in metal, hiking, rock-climbing, fencing and motorcycling.
German: In seiner Freizeit genießt Andrew Reisen und bildhauert in Metall, Wandern, Klettern, Fechten und Motorrad zu fahren.
(TCI score: 30.20)

There are other parts of a sentence that can be mis-analyzed or only partially analyzed, by either parser or transformations.

English: She read my books faster than I did *hers*.
German: Sie las meine Bücher schneller als ich *ihres* las. (TCI score: 71.10)

The reference of *hers* here was not resolved; therefore the node is still ambiguous as to singular or plural. Another one of the clean-up transformations decides on a default number assignment, but reduces the confidence score.

English: She runs a lot but I don't.
German: Sie führt viel, aber ich nicht.
(TCI score 69.80)

A lot here is resolved as the object of the verb *run*, resulting in the (incorrect) translation of a transitive verb. Transformations cannot detect this. What they can detect, however, is the fact that *viel* has

an unresolved number (singular or plural), which indicates that a resolution was missed. The same transformation as above will assign a default number and at the same time deduct 20 points from the score. Correctly, then, the confidence score for *Sie führt viel, aber ich nicht* (69.80) is lower than for *She runs every day but I don't* to German *Sie läuft jeden Tag, aber ich nicht*. (TCI score: 89.58)

The low TCI score for the final sentence in this section reflects incorrect input. The coordinator "but" hits a clean-up transformation, which indicates that it has not been processed by transformations. The reason in this case is obviously that the sentence is ungrammatical. Our TCI does not distinguish between ungrammatical input, incorrect parse and incorrect transformations.

English: Another condition involves development of a European joint venture that will collaborate in studying linguistic data but the companies to process the information separately.
German: Eine andere Bedingung schließt Entwicklung eines europäischen Joint Ventures ein, das dabei kollaborieren wird, sprachliche Daten aber die Firmen zu studieren, um die Informationen gesondert zu verarbeiten.
(TCI score: 40.76)

3.5 Dangling prepositions

Application of any of the above transformations results in either a decrease of the TCI or no change to the TCI. However, it may also be the case that application of certain transformations will *increase* our confidence. One such transformation is the *Dangling-Preposition* transformation. An unfilled slot following a preposition is penalized by 30 points in the source analysis. When, however, a particular transformation is applied, it compensates for this penalty by giving a reward of 30 points, assuming that it found the object of the preposition and improved the output.

English: Simpson was appealed to to hire an accountant.
German: An Simpson wurde appelliert, einen Buchhalter einzustellen. (TCI score: 84.00)

The German sentence needs to place preposition and prepositional object (the English subject) next to each other. Unlike English, it does not need a subject. The English parse is shown in Fig. 1 and the restructured German tree is shown in Fig. 2.
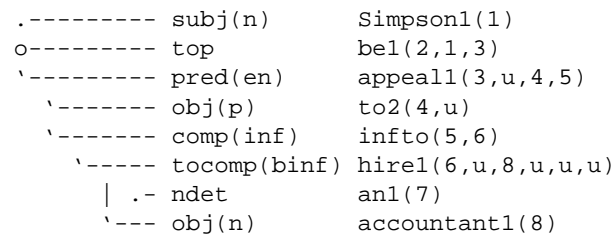
```
.--------- subj(n)        Simpson1(1)
o--------- top            be1(2,1,3)
`--------- pred(en)       appeal1(3,u,4,5)
  `------- obj(p)         to2(4,u)
  `------- comp(inf)      infto(5,6)
    `----- tocomp(binf)   hire1(6,u,8,u,u,u)
     | .-  ndet           an1(7)
     `--- obj(n)          accountant1(8)
```

Figure 1. ESG parse of "Simpson was appealed to to hire an accountant."

4 CONCLUSION

We have identified a number of potential problems when doing MT from English to German and have shown how these problems contribute to the TCI.

```
Restructured tree...
.--------- obj       an
| '------- objprep   Simpson
o--------- top       werd
'--------- pred      appellier
  | .----- punc_end  ,
  '-+----- comp
    |   .- ndet      ein
    | .--- obj       buchhalter
    '----- tocomp    (ein : stell)
    '----- punc_end  ,
```

Figure 2. Restructured German tree for "Simpson was appealed to to hire an accountant."

Source analysis has at least three ways of indicating confidence or lack thereof: a complete or incomplete parse, a parse-internal score, and the number of ambiguities. Transformations are faced with two kinds of problems that are relevant to the quality of the MT output: possibly incorrect or incomplete analysis (see Section 2.2, Section 3.1, and Section 3.3) as well as the task of changing the English parse tree sufficiently to render an acceptable German output. Where the source can state that a particular structure or word is ambiguous and therefore penalizes it, transformations can refine this scheme. This can be done either by imposing an additional penalty or by giving a reward because there is a confidence that a particular interpretation is probably correct.

How can transformations "know" whether they produce good or bad output? They may be at a disadvantage because they are several steps removed from the actual input. But they operate on a structured tree, they can examine the local syntactic, semantic, and lexical context, and they can take advantage of pattern-triggered instructions from the bilingual lexicon. Ideally, in a system with perfect parses and a complete bilingual lexicon, transformations would only have to worry about certain ambiguities and about difficulties of contrastive analysis between the source and the target language. As long as we do not have this perfect system, a tool like the TCI not only enhances the usability of an MT system, but it can also effectively be deployed as a diagnostic tool for the developers of MT.

REFERENCES

[1] Arendse Bernth, 'A confidence index for machine translation', in *Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI 99)*, pp. 120–127, Chester, England, (1999).

[2] Arendse Bernth, 'Controlling input and output of MT for greater user acceptance', in *Translating and the Computer 21*, ASLIB Conference Proceedings, p. no page numbering, London, England, (1999). ASLIB.

[3] Arendse Bernth and Michael C. McCord, 'The effect of source analysis on translation confidence', in *Envisioning Machine Translation in the Information Future, 4th Conference of the Association for Machine Translation in the Americas*, ed., John S. White, number 1934 in Springer Lecture Notes in Artificial Intelligence, pp. 89–99, Cuernavaca, Mexico, (October 2000). AMTA.

[4] Claudia Gdaniec, 'Lexical choice and syntactic generation in a transfer system: Transformations in the new LMT English-German system', in *Proceedings of AMTA-98*, pp. 408–420. Association for Machine Translation in the Americas, (1998).

[5] Michael C. McCord, 'Slot Grammars', *Computational Linguistics*, **6**, 31–43, (1980).

[6] Michael C. McCord, 'Slot Grammar: A system for simpler construction of practical natural language grammars', in *Natural Language and Logic: International Scientific Symposium, Lecture Notes in Computer Science*, ed., R. Studer, 118–145, Springer Verlag, Berlin, (1990).

[7] Michael C. McCord, 'Heuristics for broad-coverage natural language parsing', in *Proceedings of the ARPA Human Language Technology Workshop*, pp. 127–132. Morgan-Kaufmann, (1993).

[8] Michael C. McCord and Arendse Bernth, 'The LMT transformational system', in *Proceedings of AMTA-98*, pp. 344–355. Association for Machine Translation in the Americas, (1998).

[9] Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik, *A Grammar of Contemporary English*, Longman, London, 1972.

[10] Patricia M. Schmid, *Clausal Constituent Shifts: A Study in Cognitive Grammar and Machine Translation*, Ph.D. dissertation, Georgetown University, 1995.