Learning Tense Translation from Bilingual Corpora

Michael Schiehlen*
Institute for Computational Linguistics, University of Stuttgart,
Azenbergstr. 12, 70174 Stuttgart
mike@adler.ims.uni-stuttgart.de

Abstract
This paper studies and evaluates disambiguation strategies for the translation of tense between German and English, using a bilingual corpus of appointment scheduling dialogues. It describes a scheme to detect complex verb predicates based on verb form subcategorization and grammatical knowledge. The extracted verb and tense information is presented and the role of different context factors is discussed.

1 Introduction
A problem for translation is its context dependence. For every ambiguous word, the part of the context relevant for disambiguation must be identified (disambiguation strategy), and every word potentially occurring in this context must be assigned a bias for the translation decision (disambiguation information). Manual construction of disambiguation components is quite a chore. Fortunately, the task can be (partly) automated if the tables associating words with biases are learned from a corpus. Statistical approaches also support empirical evaluation of different disambiguation strategies.

The paper studies disambiguation strategies for tense translation between German and English. The experiments are based on a corpus of appointment scheduling dialogues counting 150,281 German and 154,773 English word tokens aligned in 16,857 turns. The dialogues were recorded, transcribed and translated in the German national Verbmobil project that aims to develop a tri-lingual spoken language translation system. Tense is interesting, since it occurs in nearly every sentence. Tense can be expressed on the surface lexically as well as morphosyntactically (analytic tenses).

2 Words Are Not Enough
Often, sentence meaning is not compositional but arises from combinations of words (1).

(1) a. Ich habe ihn gestern gesehen.
   I have him yesterday seen
   I saw him yesterday.

b. Ich schlage Montag vor.
   I beat Monday forward
   I suggest Monday.

c. Ich möchte mich beschweren.
   I'd like to myself weigh down
   I'd like to make a complaint.

For translation, the discontinuous words must be amalgamated into single semantic items. Single words or pairs of lemma and part of speech tag (L-POS pairs) are not appropriate. To verify this claim, we aligned the L-POS pairs of the Verbmobil corpus using the completely language-independent method of Dagan et al. (1993). Below find the results for sehen¹ in order of frequency and some frequent alignments for reflexive pronouns.

| sehen:VVFIN | be:VBZ | (aussehen) |
| 72          |       |           |
| sehen:VVFIN | do:VBP | (do-support) |
| 44          |       |           |
| sehen:VVFIN | have:VBP | (perfect) |
| 39          |       |           |
| sehen:VVFIN | see:VB |
| 35          |       |           |
| wir:PRF     | meet:VB | (sich treffen) |
| 176         |       |           |
| wir:PRF     | we:PP |
| 33          |       |           |
| sich:PRF    | spell:VBN | (sich schreiben) |
| 30          |       |           |
| ich:PRF     | forward:RP | (sich freuen auf) |
| 16          |       |           |
| wir:PRF     | agree:VB | (sich einigen) |
| 14          |       |           |
| ich:PRF     | myself:PP |
| 13          |       |           |

¹The prefix verb aus-sehen (look, be) is very frequent in the corpus, it often occurs in questions. Present sehen was frequently translated into perfect discover.
3 Partial Parsing

A full syntactic analysis of the sort of unrestricted spoken language text found in the Verb-mobil corpus is still beyond reach. Hence, we took a partial parsing approach.

3.1 Complex Verb Predicates

Both German and English exhibit complex verb predicates (CVPs), see (2). Every verb and verb particle belongs to such a CVP and there is only one CVP per clause.

(2) He would not have called me up.

The following two grammar fragments describe the relevant CVP syntax for English and German. Every auxiliary verb governs only one verb, so the CVP grammar is basically regular and implementable with finite-state devices.

\[
\begin{align*}
S \rightarrow & \ldots \ VP \ldots \\
VP \rightarrow & \text{hd:V (to) VP} \\
VP \rightarrow & \text{hd:V} \ldots \text{(Particle)}
\end{align*}
\]

\[
\begin{align*}
S \rightarrow & \ldots \text{hd:V}_\text{fin} \ldots \text{(Refi)} \ldots \text{VC} \ldots \\
S \rightarrow & \ldots \text{(Refi)} \ldots \text{VC} \ldots \\
S \rightarrow & \ldots \text{VC} \text{hd:V}_\text{fin} \ldots \text{(Refi)} \ldots \\
\text{VC} \rightarrow & \text{(VC) (zu) hd:V} \\
\text{VC} \rightarrow & \text{SeparatedVerbPrefix}
\end{align*}
\]

English CVPs are left-headed, while German CVPs are partly left-, partly right-headed.

Er wird es getan haben müssen
He will it done have must
He will have to have done it.

\[\text{\textsuperscript{2}}\text{The grammar does not handle insertion of CVPs into other CVPs and partially fronted verb complexes (3).}\]

3.2 Verb Form Subcategorization

Auxiliary verbs form a closed class. Thus, the set \(\text{sub}(v)\) of infinite verb forms for which an auxiliary verb \(v\) subcategorizes can be specified by hand. English and German auxiliary verbs govern the following verb forms.

- infinitive e.g. will
- to-infinitive (T) e.g. want
- past participle (P) e.g. get
- \(P \lor T\) e.g. have
- present participle \(\lor P \lor T\) e.g. be
- infinitive (I) e.g. müssen
- zu-infinitive (Z) e.g. scheinen
- perf.part. with \(\text{haben}\) (H) e.g. bekommen
- \(H \lor I\) e.g. werden
- \(H \lor I \lor Z\) e.g. haben
- perf.part. with \(\text{sein} \lor H \lor I \lor Z\) e.g. sein

3.3 Transducers

Two partial parsers (rather: transducers) are used to detect English and German CVPs and to translate them into predicate argument structures (verb chains). The parsers presuppose POS tagging and lemmatization. A data base associates verbs \(v\) with sets \(\text{mor}(v)\) of possible tenses or infinite verb forms.

Let \(m = |\{\text{mor}(v) : \text{Verb } v\}|\) and \(n = |\{\text{sub}(v) : \text{Verb } v\}|\). Then the English CVP parser needs \(n + 1\) states to encode which verb forms, if any, are expected by a preceding auxiliary verb. Verb particles are attached to the preceding verb. The German CVP parser is more complicated, but also more restrictive as all verbs in a verb complex (VC) must be adjacent. It operates in left-headed (S) or right-headed mode (VC). In VC-mode (i.e. inside VCs) the order of the verbs put on the output tape is reversed. In S-mode, \(n + 1\) states again record the verb form expected by a preceding finite verb \(V_{\text{fin}}\). VC-mode is entered when an infinite verb form is encountered. A state in VC-mode records the verb form expected by \(V_{\text{fin}}\) \((n + 1)\), the infinite verb form of the last verb encountered \((m)\), and the verb form expected by the VC verb, if the VC consists of only one verb \((n + 1)\). So there are \(m \ast (n + 1)^2\) states. As soon as a non-verb is encountered in VC-mode or the verb form of the previous verb does not fit the subcategorization requirements of the current verb, a test is performed to see if the verb form of the last verb

(3) Versuchen hätte ich es schon gerne wollen.
Try 'd have I it liked to
I'd have liked to try it.

1184
in VC fits the verb form required by $V_{\text{fin}}$. If it does or there is no such finite verb, one CVP has been detected. Else $V_{\text{fin}}$ forms a separate CVP. In case the VC consists of only one verb that can be interpreted as finite, the expected verb form is recorded in a new $S$-mode state. Separated verb prefixes are attached to the finite verb, first in the chain.

### 3.4 Alignment

In the CVP alignment, only 78% of the turns proved to have CVPs on both sides, only 19% had more than one CVP on some side. CVPs were further aligned by maximizing the translation probability of the full verbs (yielding 16,575 CVP pairs). To ensure correctness, turns with multiple CVPs were inspected by hand. In word alignment inside CVPs, surplus tense-bearing auxiliary verbs were aligned with a tense-marked NULL auxiliary (similar to the English auxiliary do).

### 3.5 Alignment Results

The domain biases the corpus towards the future. So only 5 out of 6 German tenses and 12 out of 16 English tenses occurred in the corpus. Both will and be going to were analysed as future, while would was taken to indicate conditional mood, hence present.

- present (15,710)
- perfect (344)
- preterite (331)
- pluperfect (49)
- future (150)
• present (12,252; progressive: 358)
• past (594; progressive: 23)
• present perfect (227; progressive: 7)
• past perfect (1; progressive: 1)
• future (1,429; progressive: 23)
• future perfect (10) • future in the past (3)

In some cases, tense was ambiguous when considered in isolation, and had to be resolved in tandem with tense translation. Ambiguous tenses on the target side were disambiguated to fit the particular disambiguation strategy.

• G present/perfect (verreist sein) (39)
• G present/past (sollte, ging) (229)
• E pres./present perfect (/lave got) (500)
• E pres./past (should, could, must) (1,218)

4 Evaluation
Formally, we define source tense and target tense as two random variables $S$ and $T$. Disambiguation strategies are modeled as functions $tr$ from source to target tense.

Precision figures give the proportion of source tense tokens $ts$ that the strategy correctly translates to target tense $tt$. Recall gives the proportion of source-target tense pairs that the strategy finds out.

$$precision_{tr}(ts, tt) = P(T = tt | S = ts, tr(ts) = tt)$$
$$recall_{tr}(ts, tt) = P(tr(ts) = tt | S = ts, T = tt)$$

Combined precision and recall values are formed by taking the sum of the frequencies in numerator and denominator for all source and target tenses. Performance was cross-validated with test sets of 10% of all CVP pairs.

4.1 Baseline
A baseline strategy assigns to every source tense the most likely target tense $tr(ts) = \arg \max_t P(t | ts)$, strategy $t$. The most likely target tenses can be read off Figures 1 and 2. Past tenses rarely denote logical past, as discussion circles around a future meeting event, they are rather used for politeness:

(5) a. Ich wollte Sie fragen, wie das aussieht.
I wanted to ask you what is on.

b. übermorgen war ich ja auf diesem Kongress in Zürich.
the day after tomorrow, I’ll be (lit: was) at this conference in Zurich.

4.2 Full Verb Information
Three more disambiguation strategies condition the choice of tense on the full verb in a CVP, viz. the source verb $(tr(ts, vs)) = \arg \max_t P(t | ts, vs)$, strategy $vs$, the target verb $(tr(t_s, vt))$, strategy $vt$, and the combination of source and target verb $(tr(ts, (vs, vt)))$, strategy $vst)$. The table below gives precision and recall values for these strategies and for the strategies obtained by smoothing (e.g. $v_st, v_s, v_t, t$ is $v_st$ smoothed first with $v_s$, then with $v_t$, and finally with $t$). Smoothing with $t$ results in identical precision and recall figures.

<table>
<thead>
<tr>
<th></th>
<th>G→E prec. recall , t</th>
<th>E→G prec. recall , t</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>.865 .865 .865</td>
<td>.957 .957 .957</td>
</tr>
<tr>
<td>$v_s$</td>
<td>.885 .854 .879</td>
<td>.970 .941 .965</td>
</tr>
<tr>
<td>$v_t$</td>
<td>.900 .876 .896</td>
<td>.973 .933 .966</td>
</tr>
<tr>
<td>$v_{st}$</td>
<td>.916 .819 .899</td>
<td>.979 .874 .965</td>
</tr>
<tr>
<td>$v_{st}, v_s, v_t$</td>
<td>.902 .892 .900</td>
<td>.970 .956 .967</td>
</tr>
<tr>
<td>$v_{st}, v_s, v_t$</td>
<td>.899 .889 .897</td>
<td>.971 .957 .967</td>
</tr>
</tbody>
</table>

We see that inclusion of verb information improves performance. Translation pairs approximate the verb semantics better than single source or target verbs. The full verb contexts of tenses can also be used for verb classifications.

Aspectual classification: The aspect of a verb often depends on its reading and thus can be better extrapolated from an aligned corpus (e.g. *I am having a drink (trinken)*). German allows punctual events in the present, English prefers present perfect (e.g. *sehen, finden, feststellen (discover, find, see); einfallen (occur, remember); treffen, erwischen, sehen (meet)).

World knowledge: In many cases perfect maps an event to its result state.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>finish</td>
<td>⇒ fertig sein</td>
<td></td>
</tr>
<tr>
<td>forget</td>
<td>⇒ nicht mehr wissen</td>
<td></td>
</tr>
<tr>
<td>denken an</td>
<td>⇒ have in mind</td>
<td></td>
</tr>
<tr>
<td>sich verabreden</td>
<td>⇒ have an appointment</td>
<td></td>
</tr>
<tr>
<td>sich vertun</td>
<td>⇒ be wrong</td>
<td></td>
</tr>
<tr>
<td>settle a question</td>
<td>⇒ (the question) is settled</td>
<td></td>
</tr>
</tbody>
</table>

4.3 Subordinating Conjunctions
Conjunctions often engender different mood.
• In conditional clauses English past tenses usually denote present tenses. Interpreting hypothetical past as present increases performance by about 0.3%.
• In subjunctive environments logical future is expressed by English simple present. The verbs vorschlagen (suggest) (in 11 out of 14 cases) and sagen (say) (2/5) force simple present on verbs that normally prefer a translation to future.

(6) I suggest that we meet on the tenth.

• Certain matrix verbs\(^3\) trigger translation of German present to English future.

4.4 Representation of Tense
Tense can not only be viewed as a single item (as sketched above, representation \(r_\text{t} \)). In compositional analyses of tense, source tense \(S\) and target tense \(T\) are decomposed into components \((S_1, \ldots, S_n)\) and \((T_1, \ldots, T_n)\). A disambiguation strategy \(tr\) is correct if \(\forall i : tr(S_i) = T_i\).

One decomposition is suggested by the encoding of tense on the surface ((present/past, 0/will/be going to/werden, 0/have/haben/sein, 0/be), representation \(r_\text{s}\)). Another widely used framework in tense analysis (Reichenbach, 1947) ((E\(\sim\)/R, R\(<\)/\(\approx\)/S, \(\pm\)progr), representation \(r_\text{r}\)) analyses English tenses as follows:

<table>
<thead>
<tr>
<th>(R\approx S)</th>
<th>(R&lt;S)</th>
<th>(R&gt;S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(\approx)R</td>
<td>present</td>
<td>past perf.</td>
</tr>
<tr>
<td>E&lt;R</td>
<td>present perf.</td>
<td>past perf.</td>
</tr>
<tr>
<td>E&gt;R</td>
<td>future</td>
<td>past perf.</td>
</tr>
</tbody>
</table>

A similar classification can be used for German except that present and perfect are analysed as ambiguous between present and future (E\(\geq\)R\(\approx\)S and E\(\leq\)R\(\geq\)S).

<table>
<thead>
<tr>
<th>repr. strat.</th>
<th>G(\rightarrow)E prec. recall , (t)</th>
<th>E(\rightarrow)G prec. recall , (t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(r_\text{t}) (t)</td>
<td>.861 .861 .861 .957 .957 .957</td>
<td></td>
</tr>
<tr>
<td>(r_\text{s}) (v_\text{s})</td>
<td>.859 .859 .859 .955 .955 .955</td>
<td></td>
</tr>
<tr>
<td>(r_\text{s}) (v_\text{t})</td>
<td>.883 .883 .876 .966 .938 .961</td>
<td></td>
</tr>
<tr>
<td>(r_\text{s}) (v_{st})</td>
<td>.894 .871 .890 .971 .933 .964</td>
<td></td>
</tr>
<tr>
<td>(r_\text{r}) (t)</td>
<td>.912 .815 .894 .978 .874 .962</td>
<td></td>
</tr>
<tr>
<td>(r_\text{r}) (v_\text{s})</td>
<td>.861 .861 .861 .964 .964 .964</td>
<td></td>
</tr>
<tr>
<td>(r_\text{r}) (v_\text{t})</td>
<td>.885 .855 .879 .973 .945 .970</td>
<td></td>
</tr>
<tr>
<td>(r_\text{r}) (v_{st})</td>
<td>.898 .875 .894 .977 .939 .972</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.915 .817 .897 .982 .878 .970</td>
<td></td>
</tr>
</tbody>
</table>

The poor performance of strategy \(r_\text{s}\) corroborates the expectation that tense disambiguation is helped by recognition of analytic tenses. Strategy \(r_\text{r}\) performs slightly worse than \(r_\text{t}\). The really hard step with Reichenbach seems to be the mapping from surface tense to abstract representation (e.g. deciding if (polite) past is mapped to logical present or past). \(r_\text{r}\) performs slightly better in E\(\rightarrow\)G, since the burden of choosing surface tense is shifted to generation.

5 Conclusion
The paper presents a way to test disambiguation strategies on real data and to measure the influence of diverse factors ranging from sentence internal context to the choice of representation. The pertaining disambiguation information learned from the corpus is put into action in the symbolic transfer component of the VerbMobil system (Dorna and Emele, 1996). The only other empirical study of tense translation (Santos, 1994) I am aware of was conducted on a manually annotated Portuguese-English corpus (48,607 English, 43,492 Portuguese word tokens and 6,334 tense translation pairs). It neither gives results for all tenses nor considers disambiguation factors. Still, it acknowledges the surprising divergence of tense across languages and argues against the widely held belief that surface tenses can be mapped directly into an interlingual representation. Although the findings reported here support this conclusion, it should be noted that a bilingual corpus can only give one of several possible translations.

References


\(^3\)ausgehen von, denken, meinen (think), hoffen (hope), schade sein (be a pity)