

Translation mining in the domain of conditionals

Jos Tellings

08/2019

- The *Time in Translation* project applies translation mining methodology in the domain of tense and aspect. The situation for conditionals is similar to that of tense: there is a lot of narrow theoretical literature based on English, and a lot of cross-linguistic variation that can inform the theory.
- Tense and aspect play a crucial role in the analysis of conditionals: the contribution of tense and aspect in the antecedent and/or consequent is thought to result in various semantic and pragmatic effects of conditionals. Some keywords: fake past tense, fake aspect, the indicative/subjunctive distinction, tense (a)symmetry, modality and tense, ...

Sections 1 to 5 provide potential topics for investigation with this methodology. Section 6 gives more details about how the translation mining methodology can be applied in the domain of conditionals.

1 One-past vs. two-past

- In English *would*-conditionals a standard contrast is made between **one-past** (one layer of past tense) and **two-past** (two layers of past tense) conditionals.

- (1) a. If John came to Mary's party next week, he would be happy.
- b. If John had come to Mary's party next week, he would have been happy.

- Three theoretical claims about the distinction:
 1. Cancellability of counterfactual implicature
 - (2) *Ippolito's generalization I* (Ippolito 2003: 147, 2013: 25-26)
Counterfactuality is cancellable in one-past subjunctive conditionals and non-future two-past subjunctive conditionals, but not in future oriented two-past subjunctive conditionals.
 2. Strength of counterfactual implicature
 - (3) *Ippolito's generalization II*
Two-past subjunctive conditionals require strong counterfactuality on *p*, whereas one-past conditionals can be strongly or weakly counterfactual on *p*.
 3. Presupposition projection
 - (4) [context: John died] (Ippolito 2006)
 - a. #If John ran the Boston marathon next spring, he would win.
 - b. If John had run the Boston marathon next spring, he would have won.

All three have been derived in some way or other from the meaning of the different tense/aspect combinations in the antecedent of the conditional.

- Other languages have similar contrasts (Dutch: modal preterit vs. past perfect (vvt), etc.). Is the one-past/two-past contrast preserved in translation?

2 Dutch *zou*

- Dutch has the same one-past/two-past contrast, but also the optional presence of modal *zou* (Nieuwint 1984; Roels et al. 2007):

- (5) a. Als ik rijk was, deed ik het.
if I rich was(PST), did(PST) I it
'If I were rich, I'd do it'

- b. Als ik rijk was, zou ik het doen.
would I it do(INF)
- c. Als ik rijk zou zijn, deed ik het.
would be(INF)
- d. Als ik rijk zou zijn, zou ik het doen.

- **Nieuwint** makes the point that the presence of *zou* can make a meaning difference:

- (6) a. Als je wat zei kreeg je straf. [✓ temporal, *counterfactual]
'If you said something, you got punished'
- b. Als je wat zou zeggen, zou je straf krijgen. [✓ counterfactual]
'If you were to say something, you would get punished'

- (7) Als je in de gevangenis zat, bezocht ik je elke dag. [*temp, ✓CF]
'If you were in prison, I'd visit you every day'

- **Nieuwint** (1984) reports on a small corpus study. **Roels et al.** (2007) use a small translation corpus of German/Dutch novels. Europarl is a much larger corpus with lots of conditionals, so we can replicate those studies.
- **Nieuwint** doesn't consider the one-past/two-past distinction, but **Roels et al.** (2007: 184) claim *zou* is less common with perfect tenses.
- Re-assessment of Nieuwint's data and proposal for *zou*, bringing it to the attention of an English audience. We can relate to more recent work on modal/evidential *zou* (e.g. **de Haan** 2001 a.o.), and make the comparison with German *würde* (**Roels et al.** 2007).

- This can lead to a wider study of modal uses of the past tense, something we have mostly ignored in the project so far. **Roels et al.** (2007) present a list of contexts in which *zou* or the modal preterite is used.

3 Tense in Dutch conditionals

- Regular tense differences between languages extend to conditionals. For example, the limited use of the Dutch simple past tense (ovt) extends to conditionals. Here is the famous Adams-sentence:

- (8) a. If Oswald didn't kill Kennedy, someone else did.
b. Als Oswald Kennedy niet **heeft vermoord**, **heeft** iemand anders dat gedaan.

- Looking at a few preliminary results from Europarl, I noticed that sometimes, English subjunctive conditionals are translated in Dutch as present tense conditionals. Are these "future less vivid" (FLV) conditionals? (**Iatridou** 2000; cf. remarks on French *imparfait* vs. *présent* in **Patard** 2011: 283):

- (9) a. If this were to be recognised in the EU, she would then be able to bring her entire family to join her.
b. Als dit in de EU wordt erkend dan kan zij vervolgens ook haar gezin hier naartoe laten komen.
- (10) a. If we were to propose that, we would of course consult the Parliament.
b. Als wij een dergelijke uitbreiding willen voorstellen, zullen wij het Parlement natuurlijk raadplegen.

- (11) a. If the directive went ahead, it would narrow the earnings gap between women and men.
- b. Als de richtlijn op deze punten wordt gerealiseerd, dan zal zij de inkomenskloof tussen mannen en vrouwen verkleinen.

4 Then

Look at the distribution of *then* cross-linguistically (Davis 1983; Iatridou 1994; von Fintel 1994; Bhatt and Pancheva 2017: §4.1; Zakkou 2017).

5 Other languages

- We want to not only compare English/Dutch/German, but include other languages too.
- Detailed work exists on tense/aspect in conditionals in Italian (Ippolito 2004), Greek (Iatridou 2000), French (Haiman and Kuteva 2002), etc. but these are mostly studies on a single language only. Our approach gives us a more direct way of compare those various languages.

6 Notes on methodology

I started with the Europarl corpus of proceedings from the European Parliament,¹ because its register makes it more likely to contain a relatively high number of counterfactual conditionals (the type I was originally interested in). Later on we may also want to look at the *Harry Potter* parallel corpus, or others.

Four steps: (1) extraction of conditionals; (2) alignment [perhaps]; (3) annotation; (4) creating semantic maps.

Step 1: extraction of conditionals from a corpus

- I wrote a very simple script (an adaptation of the PerfectExtractor script²) that extracts ‘if… would’ combinations from the English part of the Europarl corpus. This was in part because I was interested in counterfactual conditionals, but also because this is perhaps the easiest case: *would* is not that common outside of conditionals. Still there are quite a few false positives (e.g. *I do not know if we would have succeeded in the Council …*); I haven’t quantified precisely how often this happens.
 - If we want to look at conditionals more generally, both in English and cross-linguistically, this is not straightforward. A more general method for conditional extraction is thus much desired, but not easy:
 - Many (indicative) conditionals in English don’t have an overt modal like *will* (even less the word *then*). In order to extract them, a lot of syntactic parsing is necessary, which is difficult for the highly complex Europarl sentences. Random example illustrating this:
- (12) If we want individual citizens or even citizens' initiatives to become involved in this domain, we cannot confront them with time-consuming and cumbersome administrative procedures that result in their having to wait up to a year and a half before receiving a definitive response to their applications.
- In most languages, the conditional connective has several other uses (*if* as subordinating complementizer; Dutch *als* with equatives, etc.). These other uses need to be filtered out.
 - A fun project for somebody could be to look at conditional inversion (“Had I been offered the job, I would …”). This also exists in Dutch and e.g. Spanish (see Biezma 2011). However, because there is no conditional connective, they are currently not extracted.
- See for example Narayanan et al. (2009) for some computational work on extracting conditionals.

¹<http://www.statmt.org/europarl/>

²<https://github.com/UUDigitalHumanitieslab/perfectextractor>

- Europarl has the complication that the source language of the translation is not always clear. However, pre-2004 data do not use “hub languages” for translation.
- Bonus data point (from Nikki Evers’ thesis³):

- (13) a. ‘Ni skulle säkert vara stel om ni **hade suttit** på en
you would surely be sore if you had sit on a
tegelmur hela dan’, sade professor McGonagall.
brick wall all day said professor McGonagall
- b. ‘You’d be stiff if you’d been sitting on a brick wall all day’, said Professor McGonagall.

(Maybe somebody wants to look at Harry Potter at some point)

Step 2: alignment

- In the original domain of tense, alignment was crucial because the corpus was aligned at the sentence level, but alignments of individual tense forms were needed. The TimeAlign software⁴ was developed for this purpose (see Figure 1).
- Since conditionals are sentence-size units, it is not immediately clear if we need to do any alignment using TimeAlign. However, it may be useful to align *if*-clauses. For example, the annotator would mark as follows:

- (14) a. This means that **if closer cooperation were included within the Second Pillar**, it would then be necessary to find a specific way of implementing this concept in this area.
- b. Kortom, **indien bij de vaststelling van de onder de tweede pijler vallende terreinen die voor versterkte samenwerking in aanmerking komen vooruitgang wordt geboekt**, zal een specifieke

³Nikki Evers (2019). *The Perfect in Swedish Dialogue*. BA thesis.

Can be downloaded at <https://dspace.library.uu.nl/handle/1874/382895>.

⁴<https://time-in-translation.hum.uu.nl/timealign/introduction/>

formule voor de tenuitvoerlegging van dit concept in de desbetreffende context moeten worden uitgewerkt.

This can be used to automatically determine antecedent-consequent order in a fairly simple way. Furthermore, the selected (highlighted) *if*-clauses can be extracted as a subcorpus, and then be put through the existing machinery for tense alignment/annotation.

Step 3: annotation

- Decisions have to be made with respect to the categories we want to annotate for. Some first suggestions are in Table 1.

tense in antecedent, tense in consequent:	one-past, two-past, present, perfect, ...
modality:	presence of <i>zou</i> (Nl), <i>would/will</i> (En), imparfait vs. conditionnel (Fr)
type:	indicative, counterfactual, FLV, biscuit ...
order:	antecedent < consequent, consequent < antecedent

Table 1: Potential categories for annotation

- Every conditional sentence will have more than one parameter to annotate for. A parameter can either be annotated manually, or an annotation may be automatically assigned and then manually checked (I am thinking in particular of tense assignment).
- In the latter case, this can be implemented by having the annotator check some boxes:
 - The original fragment is not a conditional statement.
 - The original fragment is incorrectly marked as having past perfect in the *if*-clause.
- We have the issue of how to do multiple annotation:

1. present the annotator with a single conditional sentence and let them annotate all parameters at the same time;
2. present a single conditional sentence multiple times, each time annotating a different parameter.

I am not sure, we should figure out (i) what is most efficient for the annotator, and (ii) what is easiest to implement in the annotation software.

Step 4: semantic maps and distance function

- Old situation:

data points:	individual tense forms
context:	sentence in which they occur

Let's call the standard distance function between tuples 'd'. Example:

$$d(\langle \text{Perf}, \text{vtt}, \text{Perfekt}, \text{Futur} \rangle, \langle \text{Perf}, \text{vtt}, \text{Prät}, \text{Imparfait} \rangle) = 2/4 = .5$$

- Conditionals:

data points:	conditional sentences
context:	preceding/following sentences

Each point contains more than one annotation label. For example an English conditional could be annotated as [past, past, would] (the order is [ANT_TENSE, CONS_TENSE, MOD]). A tuple of translations may then look like $\langle [\text{Past}, \text{Past}, \text{would}], [\text{ott}, \text{vtt}, \text{zou/zou}], [\text{imparf}, \text{conditionnel}, ??^5] \rangle$ for example.

- We can now apply our d-function coordinatewise. To illustrate, let $t_1 = \langle [\text{Past}, \text{Past}, \text{would}], [\text{ott}, \text{vtt}, \text{zou/zou}] \rangle$, and

⁵The 'modal' parameter is tricky because this is realized in so many different ways: modal auxiliary in consequent (En), modal auxiliary in consequent and/or antecedent (Dutch), conditionnel mood in French, ... We'll have to think how to deal with this.

$t_2 = \langle [\text{Past}, \text{Perf}, \text{would}], [\text{ott}, \text{vtt}, \text{zou/-}] \rangle$. Then:

$$\begin{aligned} d_1(t_1, t_2) &= d(\langle \text{Past}, \text{ott} \rangle, \langle \text{Past}, \text{ott} \rangle) = 0; \\ d_2(t_1, t_2) &= d(\langle \text{Past}, \text{vtt} \rangle, \langle \text{Perf}, \text{vtt} \rangle) = 1/2 = .5; \\ d_3(t_1, t_2) &= d(\langle \text{would}, \text{zou/zou} \rangle, \langle \text{would}, \text{zou/-} \rangle) = (0 + .5)/2 = .25; \end{aligned}$$

Informally: d_n = standard d-function when you only take the n -th coordinate into consideration.

- This effectively gives access to the component annotations of the conditionals, so d_1 can be used to create a map of the antecedent tense in the conditionals, and compare it to earlier maps (in view of section 3 above).
- On the other hand, a combined distance can now be defined. For example, the mean of d_1, d_2, d_3 , which would give .25 in the example above.
- Ideally (I don't know if this is practically possible at all) you can select the distance function you want for your semantic map in the web interface.

Issue Sometimes a conditional in one language is not translated as a conditional in another language, although this doesn't seem to be very common. Example:

- (15)
- a. If the Darfur Peace Agreement were implemented – and barely any progress has been made on this – it would have an immediate impact on the lives of some six million Darfurians.
 - b. De tenuitvoerlegging van het Vredesakkoord van Darfoer, waarin vrijwel geen vooruitgang is geboekt, heeft directe gevolgen voor het leven van ongeveer zes miljoen mensen uit Darfoer.

This would not constitute a data point in Dutch under the approach given above.

References

- Bhatt, R., & Pancheva, R. (2017). Conditionals. In M. Everaert & H. van Riemsdijk (eds.), *The Wiley Blackwell Companion to Syntax, Second Edition*. Wiley Online Library.

- Biezma, M. (2011). Conditional inversion and GIVENNESS. In *Proceedings of SALT 21* (pp. 552–571).
- Davis, W. (1983). Weak and strong conditionals. *Pacific Philosophical Quarterly*, 64(1), 57–71.
- von Fintel, K. (1994). *Restrictions on Quantifier Domains*. Doctoral dissertation, University of Massachusetts, Amherst.
- de Haan, F. (2001). The relation between modality and evidentiality. *Linguistische Berichte*, 9, 201–216.
- Haiman, J., & Kuteva, T. (2002). The symmetry of counterfactuals. In J. Bybee & M. Noonan (eds.), *Complex sentences in grammar and discourse: Essays in honor of Sandra A. Thompson*, pp. 101–124. John Benjamins.
- Iatridou, S. (1994). On the contribution of conditional *then*. *Natural Language Semantics*, 2, 171–199.
- Iatridou, S. (2000). The Grammatical Ingredients of Counterfactuality. *Linguistic Inquiry*, 31(2), 231–270.
- Ippolito, M. (2003). Presuppositions and implicatures in counterfactuals. *Natural Language Semantics*, 11, 145–186.
- Ippolito, M. (2004). Imperfect modality. In J. Guéron & J. Lecarme (eds.), *The syntax of time*, pp. 359–387. Cambridge, MA: MIT Press.
- Ippolito, M. (2006). Semantic Composition and Presupposition Projection in Subjunctive Conditionals. *Linguistics and Philosophy*, 29, 631–672.
- Ippolito, M. (2013). *Subjunctive conditionals*. Cambridge, MA: MIT Press.
- Narayanan, R., Liu, B., & Choudhary, A. (2009). Sentiment analysis of conditional sentences. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1* (pp. 180–189). Association for Computational Linguistics.
- Nieuwint, P. (1984). Werkwoordtijden in Nederlandse counterfactuals [Verb tenses in Dutch counterfactuals]. *De Nieuwe Taalgids*, 77, 542–555.
- Patard, A. (2011). The epistemic uses of the English simple past and French *imparfait*. In A. Patard & F. Brisard (eds.), *Cognitive approaches to tense, aspect, and epistemic modality*, pp. 279–310. John Benjamins.
- Roels, L., Mortelmans, T., & van der Auwera, J. (2007). Dutch equivalents of the German past conjunctive: *zou* + infinitive and the modal preterit. In L. de Saussure, J. Moeschler, & G. Puskás (eds.), *Tense, Mood and Aspect*, pp. 177–196. Brill Rodopi.
- Zakkou, J. (2017). Biscuit Conditionals and Prohibited ‘Then’. *Thought: A Journal of Philosophy*, 6(2), 84–92.

Annotation

English (original)

ep-11-01-17-019.xml - 122

If the directive went ahead, it would narrow the earnings gap between women and men.

Dutch (translated)

ep-11-01-17-019.xml - 134

Als de richtlijn op deze punten wordt gerealiseerd, dan zal zij de inkomenenkloof tussen mannen en vrouwen verkleinen.

- The selected words in the original fragment do not form an instance of (a/an) *One*
 This is a correct translation of the original fragment

Tense

Other label

Comments

Figure 1: Conditionals in TimeAlign (showing the old version for tense annotation only)