Conditionals in Translation

Towards Translation Mining in a compositional setting

Jos Tellings

Utrecht University ✉ j.tellings@gmail.com

30 Oct 2020

Conditionals: motivation

Conditionals are *if-then* sentences, combining an antecedent clause (*if*-clause) with a consequent clause (main clause):

   (1)   If this workshop had been held offline, it would have been more fun.

- Conditionals, like tense, have been studied widely from a formal perspective, but display a lot of cross-linguistic variation.
- Conditionals are compositional structures: tense, aspect, and modality work together to produce semantic and pragmatic effects.
- We can compare the variation of tense in conditionals with what we found for tense variation more generally.

Data: Europarl corpus

- I used the Europarl parallel corpus of proceedings of the European Parliament.
- It contains a high number of conditional utterances, presumably due to its more formal register.
- I used a script to extract English *would*-conditionals and their translations. This yields a large dataset (∼21k *would*-conditionals).
- For today's study I extracted 1000 English *would*-conditionals, together with their Dutch translations.
  (for comparision: the Camus-study has 348 verb phrases (van der Klis et al. 2020), the Harry Potter study 800 (van der Klis et al. 2019))
- Angelica Hill analyzed 250 English *would*-conditionals with their Spanish translations.

Annotation

- The data were annotated for multiple features.
  English-Dutch dataset:
    1. Antecedent tense
    2. Consequent tense
    3. modality (for English: modal verb in the consequent (*would*, *could*, *will*, etc.); for Dutch: modal verb in antecedent and consequent)
- English-Spanish dataset:
    1. Antecedent tense
    2. Consequent tense
    3. Spanish complementizer (*si*, *que*)
    4. clause order

Background: conditionals in English

- For *would*-conditionals, a distinction is made between one-past and two-past conditionals (example from Arregui 2007):

  (2) a. If your plants died next week, I would be very upset.

  b. If your plants had died next week, I would have been very upset.

- A future-oriented one past conditional is also known as a *future less vivid conditional* (FLV, Iatridou 2000). The simple past can also be used in a past-oriented indicative conditional:

  (3) a. [How should John travel to Paris tomorrow?]
      If John took the train, he would be late. [FLV]

  b. [John always pays his fare]
      If John took the train, he bought a ticket. [past indic]

Background: past reference in Dutch

- The default tense to refer to past events in Dutch is the PERFECT, called the **vtt**. The PAST, called the **ovt**, is infelicitous in these out-of-the-blue contexts:

  (4) a. Jan heeft de trein genomen.   [vtt]
           John has   the train taken

      b.# Jan nam de trein.   [ovt]
           John took the train

- The ovt is licensed in certain configurations (with stative verbs, in narrative sequences, etc.), and may have a modal interpretation in conditionals.

Background: *zou*

- The ovt can get a modal interpretation, but Dutch also has the optional presence of modal *zou* in antecedent, consequent, or both (Nieuwint 1984; Roels et al. 2007):

  (5) a. Als ik rijk was (**ovt**), deed (**ovt**) ik het.
          if   I   rich was,   did   I   it
          'If I were rich, I would do it'

       b. Als ik rijk was (**ovt**), zou   ik het doen (**inf**).
                                        would I   it   do

       c. Als ik rijk zou   zijn (**inf**), deed (**ovt**) ik het.
                     would be

       d. Als ik rijk zou zijn (**inf**), zou ik het doen (**inf**).

- All four combinations are possible in principle, but there are contextual restrictions and potential meaning differences (Nieuwint 1984, Haeseryn et al. 1997, §28.3.3.2.2).

Prediction:

- English simple pasts in antecedents that have a modal reading are translated by a modal ovt or *zou* + infinitive in Dutch;
- English simple pasts in antecedents that have a past reading (e.g. in past indicatives) should replicate our earlier findings for Dutch past reference (vtt vs. ovt).

Example: In Dutch, the FLV is translated with a modal ovt or *zou*+infinitive. For the past indicative, we use the vtt, as is typical for past reference:

(6) a. Als Jan de trein nam [ovt] / zou nemen, zou hij te laat zijn.
      'If John took the train, he would be late' [FLV]

   b. Als Jan de trein heeft genomen [vtt], heeft hij een kaartje gekocht.
      'If John took the train, he bought a ticket' [past indicative]

Demonstration of interface

Three things we learn from the maps:

1. A rather large proportion (23%) of the English conditionals is translated by a non-conditional construction in Dutch (nominalizations, infinitival clauses, factive clauses, etc.).

2. The Dutch tense distribution does not mirror the English one-past / two-past dichotomy.

3. The distribution of modal *zou*.

Tense distribution in Dutch conditionals

- The Past Perfect in the antecedent is quite stable, the Simple Past is not.
- A large number of Simple Past antecedents (33%) is translated by the **ott** (the PRESENT).
- Many of the datapoints in the ott cluster are future oriented conditionals (FLV/FNV):

  (7) If the amendment were to be adopted, all fertiliser – natural fertiliser – within the EU would have to be marked.
  Als het amendement aangenomen wordt [ott], zou onder andere alle kunstmest en mest in de Europese Unie gemerkt moeten worden.

  The use of the ott in these cases may relate to the fact that the Dutch ott can refer to the future.

- The English *be to* construction (modal/future, Declerck 2010) is quite common as well, which has no direct Dutch counterpart.

- Real counterfactual (Dutch ovt):

  (8)   We would still have to make savings if there were no
        monetary union.
        We zouden ook moeten bezuinigen als er geen monetaire
        unie was [ovt].

- Past reference (Dutch vtt):

  (9)   If the House voted last Monday to hold this debate today,
        then we must abide by that vote.
        Als dit Huis gewenst heeft [vtt] afgelopen maandag bij een
        stemming dat we dit debat nog zouden voeren vandaag, dan
        moeten wij ook dat respect hier met elkaar kunnen opbrengen

Methodological conclusions

- Using a balanced data set is important:
  - also extract Dutch conditionals, along with their English translations
  - only extracting *would*-conditionals yields a restricted dataset
- MDS maps that display several semantic features tend to be difficult to read.
  - The maps have a high Stress value (badness-of-fit measure).
  - At the moment, only two languages are included, but the power of Translation Mining lies in having several languages.
- In the current maps, the similarity between translation tuples (the distance function) is no different from the one used for single-feature MDS maps.
  - Outside of linguistics multi-mode MDS has been used: MDS with multiple dissimilarity values for each pair of objects (de Leeuw and Mair 2009).

References

Declerck, R. (2010). "Future time reference expressed by *be to* in Present-day English".
  In: *English Language & Linguistics* 14.2, pp. 271–291.

Declerck, R. and S. Reed (2001). *Conditionals. A Comprehensive Empirical Analysis*.
  Berlin / New York: Mouton de Gruyter.

Haeseryn, W. et al., eds. (1997). *Algemene Nederlandse Spraakkunst*.

Iatridou, S. (2000). "The Grammatical Ingredients of Counterfactuality". In: *Linguistic
  Inquiry* 31.2, pp. 231–270.

Van der Klis, M., B. Le Bruyn, and H. de Swart (2019). "Ways of temporal reference in
  written and spoken discourse". Ms., Utrecht University.

Van der Klis, M., B. Le Bruyn, and H. de Swart (2020). "A multilingual corpus study of
  the competition between PAST and PERFECT in narrative discourse". Ms., under
  review.

de Leeuw, J. and P. Mair (2009). "Multidimensional scaling using majorization:
  SMACOF in R". In: *Journal of Statistical Software* 31.3.

Nieuwint, P. (1984). "Werkwoordstijden in Nederlandse counterfactuals [Verb tenses in
  Dutch counterfactuals]". In: *De Nieuwe Taalgids* 77, pp. 542–555.

Roels, L., T. Mortelmans, and J. van der Auwera (2007). "Dutch equivalents of the
  German past conjunctive: *zou* + infinitive and the modal preterit". In: *Tense, Mood
  and Aspect*. Ed. by L. de Saussure, J. Moeschler, and G. Puskás. Brill Rodopi,
  pp. 177–196.

Non-conditional translations

- High number of non-conditional translations: 770/1000 are translated by a conditional construction in Dutch. We find nominalizations, infinitival clauses, factive clauses, . . .
- Example: infinitival clause

    (10)   a.   It would also be a good idea if we were to give a little consideration to our own history.

            b.

Het	zou	ook	goed	zijn	om	eens	naar	onze	eigen
it	would	also	good	be	to	once	to	our	own

geschiedenis	te	kijken.
history		to look

            lit. 'It would be good to look at our own history'

   ⇒ seminominal conditionals (Declerck and Reed 2001, §11.11)

- English non-conditional translations of Dutch conditionals not part of the dataset.

Distribution of *zou*

- The distribution of *zou* in antecedent and consequent, split by tense, is as follows:

	cons *zou*		cons X		
ant *zou*	vt	25	vvt	4	29
	inf	128	inf	2	130
ant X	vvt	126	vvt	138	264
	inf	52	ovt	3	55
		331		147	478

- 'IF *zou*, THEN X' combination is very uncommon, replicating observations in Nieuwint (1984) and Roels et al. (2007).
- Asymmetry: (v)vt conditionals disprefer *zou* in the antecedent; ovt/inf conditionals prefer *zou*, especially in the consequent.

- Roels et al. (2007) report data from the CGN (Corpus of Spoken Dutch):

	cons *zou*		cons X		
ant *zou*	vt	16	vvt	2	18
	inf	198	inf	19	217
ant X	vvt	47	vvt	33	80
	inf	94	ovt	21	115
		355		75	430

(a) CGN

	cons *zou*		cons X		
ant *zou*	vt	25	vvt	4	29
	inf	128	inf	2	130
ant X	vvt	126	vvt	138	264
	inf	52	ovt	3	55
		331		147	478

(b) Europarl

Table: Distribution of *zou* in conditionals

- The two datasets are not directly comparable, but some of the relative differences are striking.
- Further qualitative analysis can connect the corpus results with theoretical status of *zou* (Nieuwint's 1984 irrealis vs. potentialis difference).

Spanish data

Angelica Hill's work on Spanish

- Annotation of Spanish translations of ca. 250 English two-past *would*-conditionals from Europarl.

- Antecedent tense

subjuntivo pretérito pluscuamperfecto	179	85%
subjuntivo pretérito imperfecto	15	7%
other	16	8%

  Consequent tense

condicional (compuesto)	169	78%
subjuntivo pretérito pluscuamperfecto	39	18%
other	8	4%

(11)   Si hubiésemos sido [PCP] un poco más rápidos, esta directiva
       habría estado [COND] lista con bastante anterioridad al llamado
       Protocolo de Goteburgo.
       If we had been a little quicker off the mark, the present directive
       would have been on the table – signed, sealed and delivered –
       long before the so-called Gothenburg Protocol.

(12)   Si el Grupo Liberal no hubiera abogado [PCP] por un
       tratamiento rápido en la Comisión de Asuntos Jurídicos, no
       hubiésemos tenido [PCP] un informe el 3 de diciembre.
       If the Liberals had not forced the pace of the discussions in the
       Legal Affairs Committee, there would have been no report on 3
       December.