

Forecat

Translation technologies are being increasingly used to assist human translators. Within this context, the objective of interactive translation prediction (ITP) tools is to assist human translators in the translation of texts for dissemination by making context-based computer-generated suggestions as they type. Most works in the field of ITP have solely used specifically-adapted statistical machine translation (SMT) systems to obtain the suggestions. On the contrary, the *resource-agnostic* approach considered for the tool explores how non-adapted *black-box* bilingual resources of any kind (a machine translation system, a translation memory, a bilingual dictionary, etc.) can be accommodated into an interoperable ITP framework.

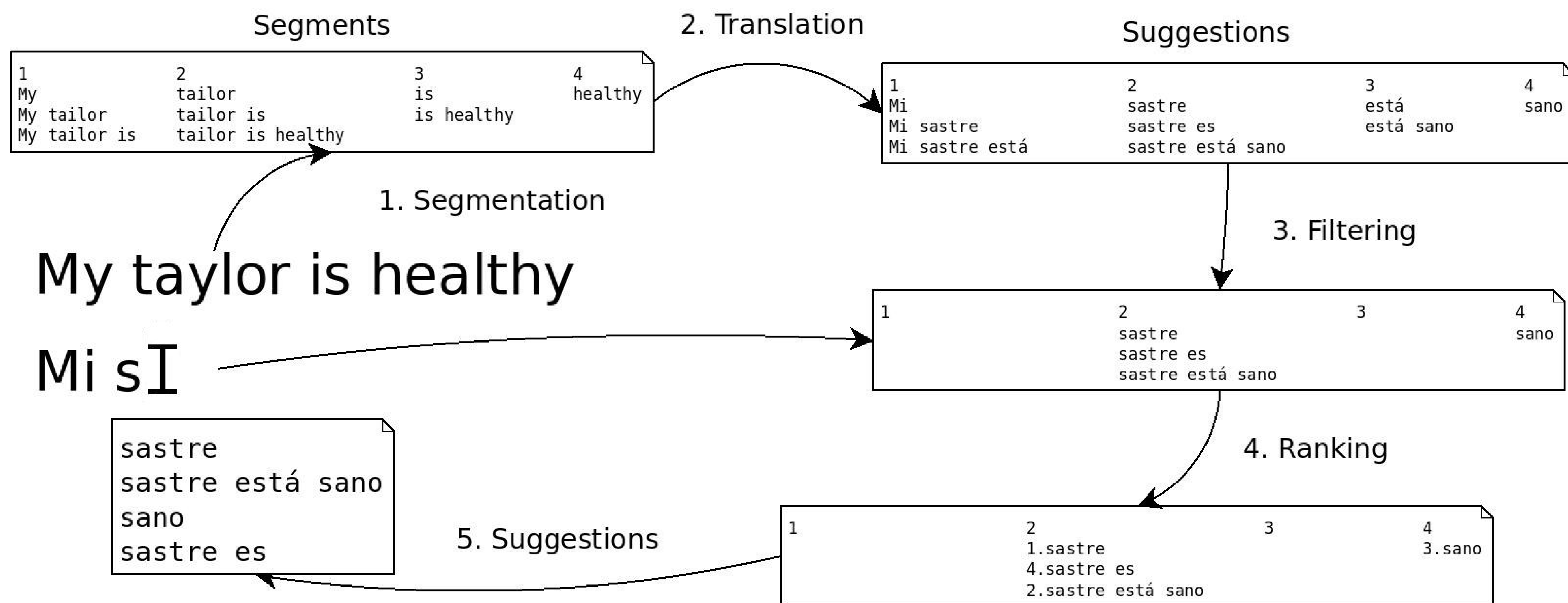


Figure: Workflow of the tool.

A resource-agnostic interactive translation prediction approach

Potential suggestions: split in all possible overlapping sentences up to length L and translate it using any bilingual resource.

Compatible suggestions: translations that have the last (partially written) word as prefix are considered as compatible.

Ranking suggestions: only the suggestion closer to the position currently being translated are prioritized, up to a fixed value. Only the longest and shortest suggestion of each position is offered, filling with the rest if there are not enough.

Usage of the tool

Forecat is an open-source web-based tool that we have implemented to demonstrate the validity of our ITP approach and incorporate its use in real-life applications. Forecat can be used in three different ways:

It can be used as a simple web application for computer-assisted translation. Users freely type the translation of the source sentence, and are offered suggestions *on the fly* based on the current prefix; users may accept these suggestions or ignore them and continue typing.

The usage of *Google Web Toolkit* lets to balance the load in three different ways: with the server doing all the computations, only translating or with the client doing all the work.

It can be deployed as a set of *web services* with an *application programming interface* (API) that provides the basic functionalities for integrating our ITP method in any third-party application via *JSON*.

It can be included in any webpage as a *web component*. Web components comply with a number of standards whose objective is to enable fully encapsulated and reusable components for the web, letting you include advanced functionalities in a simple way.

Experimental results

	en-es	es-en	en-es	cs-en
ep	0.70	0.53	0.78	0.69
ep+nc	0.62	0.52	0.75	0.66
nc	0.62	0.52	0.76	0.64

The performance of our system is measured by using the *keystroke ratio*; *ep* stands for the system trained and tuned with Europarl (out-of-domain); *nc* for the one trained and tuned with News Commentary (in-domain); and *ep+nc* for the one trained with Europarl but tuned with News Commentary. The *test set* consisted of sentences extracted from the News Commentary corpus. For the purposes of comparison, the rule-based MT system Apertium scored 0.76 for English-Spanish and 0.70 for Spanish-English.

Web Component

```
<!DOCTYPE html>
<html>
<head>
  <script src="bower_components/platform/platform.js"></script>
  <link rel="import" href="elements/translation-box.html">
</head>
<body>
  <translation-box id="itp"></translation-box>
  <script>
    var component = document.querySelector('#itp');
    component.addEventListener('languagesReady', function (e) {
      if (contains(e.detail, "en-es")) {
        component.pair= "en-es";
        component.sourceText= "My tailor is healthy.";
      }
    });
  </script>
</body>
```

Figure: HTML code for including Forecat web component.

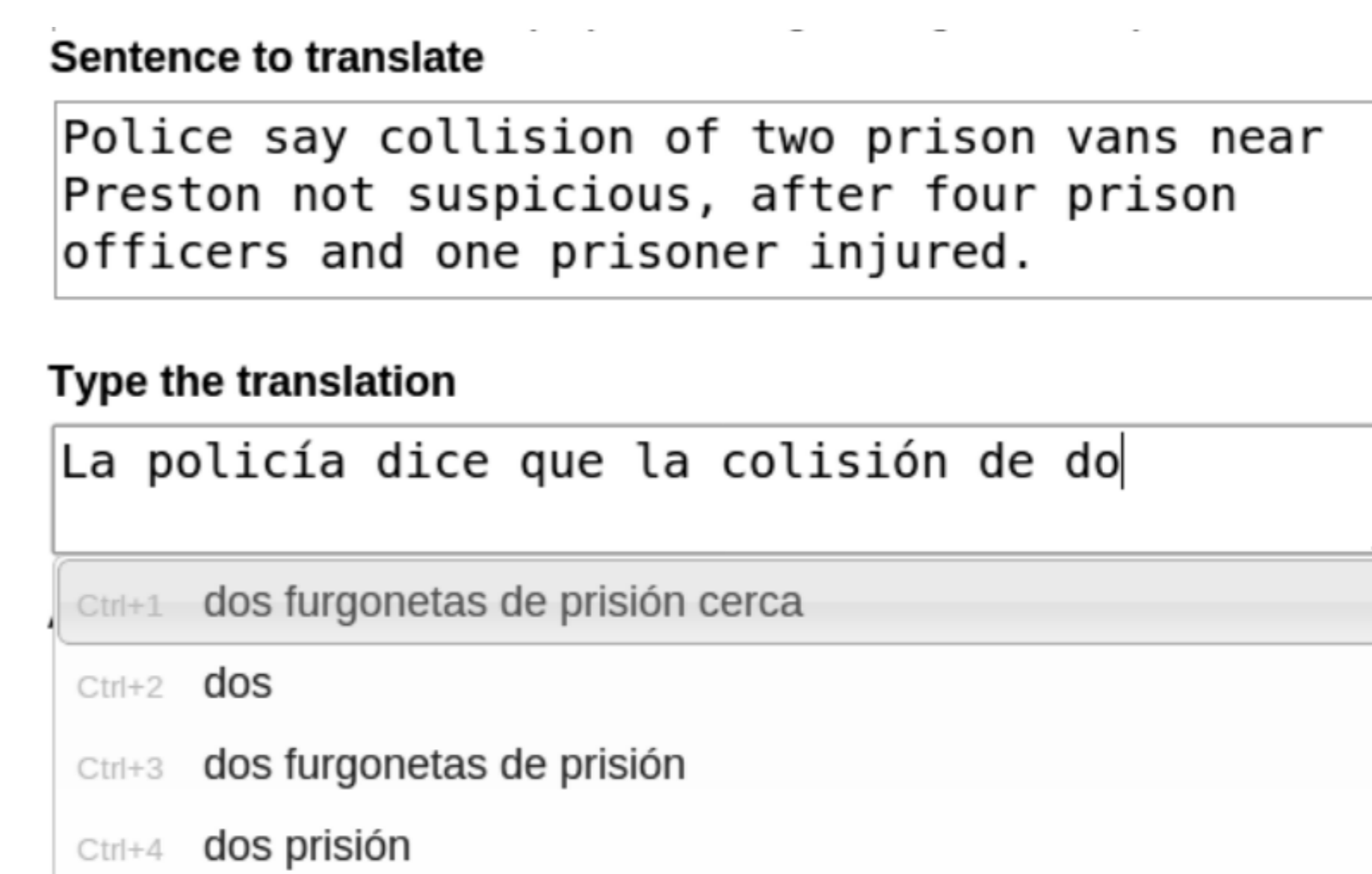


Figure: How the web component looks.

Acknowledgments

This work has been partly funded by the Spanish Ministerio de Economía y Competitividad through project TIN2012-32615.