



# **D4.8 Pilot IV: First Report**

Author(s): Katharina Allgaier,  
Susana Veríssimo, Matthias  
Orlikowski, Matthias Hartung  
(SEM)

Date: December 31st, 2020



**H2020-ICT-29b**

**Grant Agreement No. 825182**

Prêt-à-LLOD - Ready-to-use Multilingual  
Linked Language Data for Knowledge  
Services across Sectors

*D4.8 Pilot IV First Report*

Deliverable Number: 4.8

Dissemination Level: Public

Delivery Date: M24

Version: v1

Author(s): Katharina Allgaier, Susana Veríssimo, Matthias Orlikowski, Matthias Hartung (SEM)

**Document History**

<b>Version Date</b>	<b>Changes</b>	<b>Authors</b>
31/10/2020	Content structure, First draft	Matthias Hartung
17/12/2020	Draft version for internal review	Katharina Allgaier, Susana Veríssimo, Matthias Orlikowski, Matthias Hartung
21/12/2020	Internal review	John McCrae (NUIG), Jorge Gracia (UNIZAR)
22/12/2020	Final version	Matthias Hartung



# Table of Contents

<b>1 Background and Motivation</b>	<b>4</b>
<b>2 LLOD Processing Pipeline</b>	<b>5</b>
<b>3 LTTL Architecture</b>	<b>6</b>
3.1 Transfer Learning Framework	6
3.2 Software Architecture	7
<b>4 Language Resources</b>	<b>8</b>
4.1 Monolingual Word Embeddings	8
4.2 Bilingual Translation Dictionaries	9
4.2.1. Single-source Lexica	9
4.2.2. Extended Lexica	10
4.3 Cross-lingual Lexicon Induction	11
<b>5 Experiments and Results</b>	<b>13</b>
5.1 Cross-lingual Sentiment Analysis	13
5.1.1. Medical Experts' Transcripts	14
5.1.2. Lexical resources and experiment configurations	14
5.1.3. Results	15
5.2 Cross-lingual Concept Detection	15
5.2.1 Available monolingual model and generation of test set labels	16
5.2.2 Baseline Models	16
5.2.3 Experiment 1: Medical Experts' Transcripts	17
5.2.4 Experiment 2: Social Media Data Experiments	19
<b>6 Summary and Conclusions</b>	<b>24</b>
<b>References</b>	<b>24</b>
<b>Appendix A</b>	<b>25</b>



# Language- and Task-independent Transfer Learning for Multilingual Text Analytics in the Pharmaceutical Domain based on Linguistic Linked Open Data Resources and Workflows

## 1 Background and Motivation

Semalytix provides solutions to customers in the pharmaceutical industry in order to foster patient centricity in medical drug development through text analytics on real-world data. According to the FDA, patient-centric drug development is “a systematic approach to help ensure that patients’ experiences, perspectives, needs, and priorities are captured and meaningfully incorporated into drug development and evaluation. As experts in what it is like to live with their condition, patients are uniquely positioned to inform the understanding of the therapeutic context for drug development and evaluation”. Crucially, the aforementioned definition refers to real-world patient experience (“what it is like to *live* with their condition”) rather than clinical outcomes that are regularly measured and evaluated under laboratory conditions in clinical trials. Real-world insights of this kind, comprising self-reported disease burdens, treatment experience and unmet needs of particular patient populations, are increasingly gaining importance in regulatory approval and value assessment procedures of new drug products. Semalytix provides access to these insights via their Pharos Pharma Analytics platform, based on text analytics from large volumes of patient-reported narratives that are gathered in social media. As an additional source of evidence, health outcomes reported through medical experts are used to complement the patients’ perspective.

Given that pharmaceutical companies usually operate in multiple markets across the globe and approximately 40% of the global pharmaceutical annual revenue is generated in regions with native languages other than English, it is obvious that a strong demand for text analytics in multiple languages arises from this problem setting. However, the technical challenge of providing and maintaining an analytics stack for the entirety of involved NLP tasks (entity tagging, concept detection, sentiment analysis, among various others) is too expensive to be addressed from scratch with dedicated language-specific models for every single language of interest. In addition, the language adaptation challenge is aggravated by an inherent domain adaptation problem due to the specifics of pharmaceutical/biomedical domain plus text genre effects (patient-generated social media narratives vs. medical experts’ reports from CRM data).

Therefore, in this pilot project, we are working towards a language- and task-independent transfer learning (LTTL) architecture for cross-lingual projection of existing NLP models that were monolingually trained on domain-specific English data for which sufficient resources are available. LTTL avoids manual data labeling efforts in the target language by relying on a data processing pipeline that is based on Linguistic Linked Open Data (LLOD), thus

capitalizing on existing linguistic resources (bilingual translation dictionaries, monolingual word embeddings) from the openly accessible LLOD cloud and dedicated LLOD workflows developed in the Prêt-à-LLOD project. We demonstrate that LTTL can be effectively rolled out to a variety of NLP tasks (sentiment analysis and concept detection) in different languages (Spanish, French) and across text genres (social media patient narratives, medical experts reports from CRM data). In task- and language-specific experiments, LTTL is evaluated against sequential machine translation pipelines. From the results obtained, we can conclude that LTTL provides a very reasonable trade-off on the cost-effectiveness spectrum, outperforming machine translation baselines in many evaluation scenarios.

The structure of this report is as follows: In Section 2, we describe the LLOD processing pipeline that underlies LTTL in order to fuel the learning framework with the required linguistic resources. Section 3 provides details about the technical underpinnings of the LTTL core, both in terms of learning framework and software architecture. An overview of the concrete linguistic resources (along with information about their most important characteristics and provenance) used in the experiments reported in this document is given in Section 4. Section 5 presents two experiments on cross-lingual transfer of sentiment analysis and cross-lingual concept detection. Section 6 covers conclusions and next steps, including an alignment of the achieved progress in the pilot so far, compared to the original work plan.

## 2 LLOD Processing Pipeline

In order to avoid manual data labeling efforts in the target language, LTTL capitalizes on existing LLOD resources (bilingual translation dictionaries, monolingual word embeddings). Figure 1 presents the data processing pipeline that has been developed in the Prêt-à-LLOD project in order to transform existing lexical resources into RDF and publish them as LLOD which is subsequently consumed by LTTL in order to induce task-specific machine learning models in a target language of interest that are needed in the Pharos NLP stack.

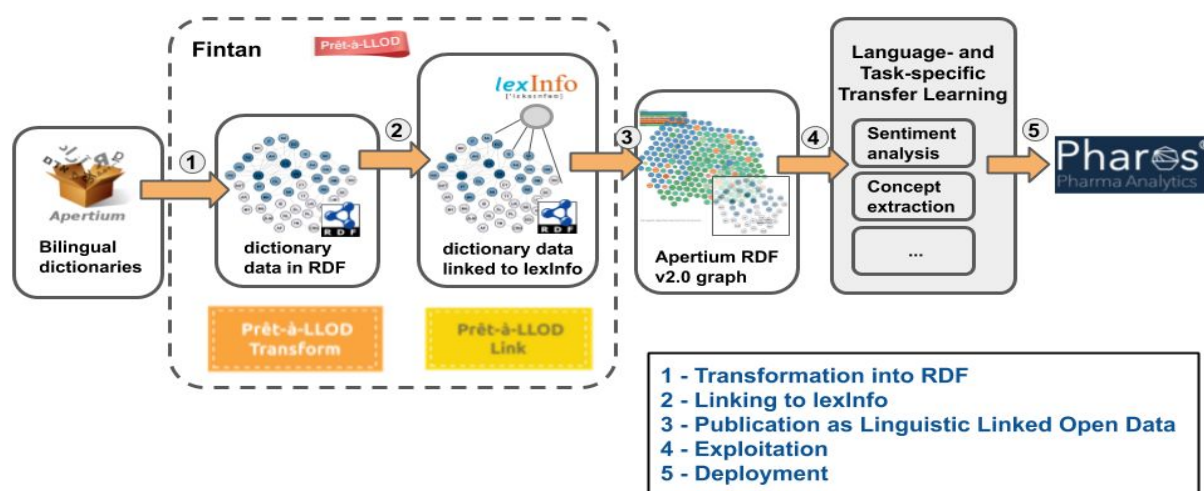


Figure 1: Overview of LLOD Processing Pipeline (slightly adapted from Gracia et al., 2020)

For a more detailed description of the LLOD pipeline and the individual processing steps involved, we refer to Gracia et al. (2020). Due to its genericity, this workflow has the potential to serve as a strong enabler of growing the LLOD cloud over time both in terms of data volume and richness. In combination with a language- and task-independent

transfer learning framework such as LTTL, we consider it as a highly versatile catalyzer for rapid induction of multilingual text analytics components.

## 3 LTTL Architecture

### 3.1 Transfer Learning Framework

We approach language- and task-independent transfer learning (LTTL) by applying supervised cross-lingual projection methods to generate bilingual word representations (Sogaard et al 2019). We in particular draw on a neural network architecture to learn *Bilingual Sentiment Embeddings* (Barnes et al. 2018) which encode information relevant to sentiment analysis in a shared bilingual word embedding space. Based on the assumption that sentiment classification is not essentially different from other classification tasks, we adopt BLSE as a general framework. This allows us to learn bilingual task-informed embedding spaces for any task which can be expressed as text classification. For this, LTTL requires 1) monolingual word embeddings in both the source and target language, 2) ground-truth annotations in the source language, and 3) a bilingual dictionary that maps tokens from the source language to their translations in the target language. Annotations in the target language are required for evaluation only.

During training (Figure 2), for each document in the source-language annotated corpus we look up the (source-language) word embeddings of its tokens, average them to represent the document and project this document vector  $a_s$  using a matrix  $M_s$ . The resulting vector  $z_s$  is then passed to a softmax layer to derive the predicted label. Based on minimizing a cross-entropy loss between the predicted and the annotated labels,  $M_s$  and the parameters of the softmax layer are learnt to produce better sentiment predictions. Simultaneously, during each training loop, for every pair in the bilingual dictionary, we look up their word embeddings in the respective monolingual embedding space and project them using matrices  $M_s$  (source language) and  $M_t$  (target language) respectively. We optimize both matrices to minimize the Euclidean distance between the projected embeddings, so that the projections from the target language are as close as possible to the projections from the source language.

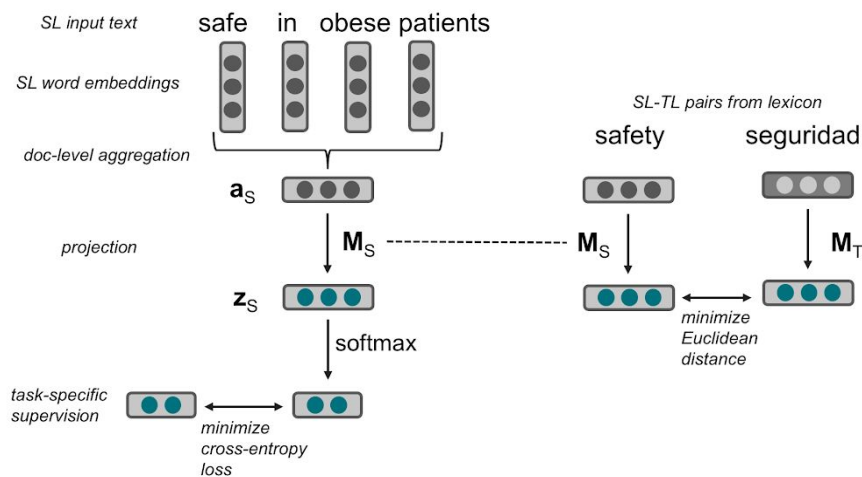


Figure 2: Training LTTL on a source-language (SL) annotated corpus and a source-language to target-language (TL) bilingual lexicon using TL and SL word embeddings to represent individual tokens

When using a trained LTTL model to classify a target-language document (Figure 3), we apply the same steps as during training based on target-language embeddings (embedding lookup, averaging, projection, prediction using the softmax layer). The projection step, however, is calculated using the matrix  $M_T$  which was optimized to project target-language embeddings close to the projections from the task-informed, source-language projection matrix  $M_S$ . Thus, the resulting vectors are expected to result in equally appropriate predictions when passed to the softmax layer learnt with the source-language annotations.

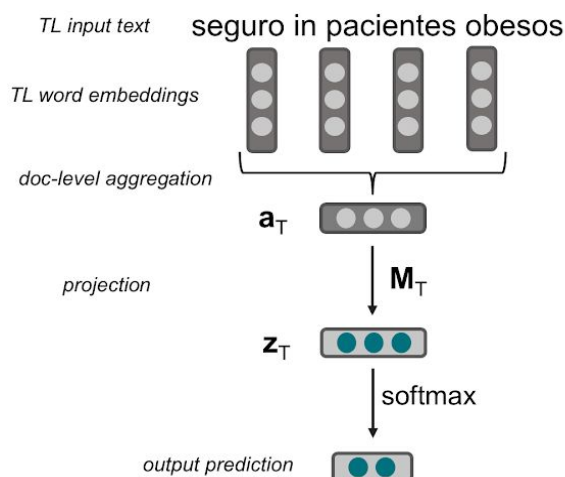


Figure 3: Predictions with LTTL on target-language (TL) text using TL word embeddings to represent individual tokens

## 3.2 Software Architecture

As described in the previous section, to train a bilingual model the LTTL framework requires bilingual lexical resources, monolingual pre-trained embeddings and annotated task-specific data in the source language. The software architecture that integrates these different resources into the learning framework is shown in Figure 4.

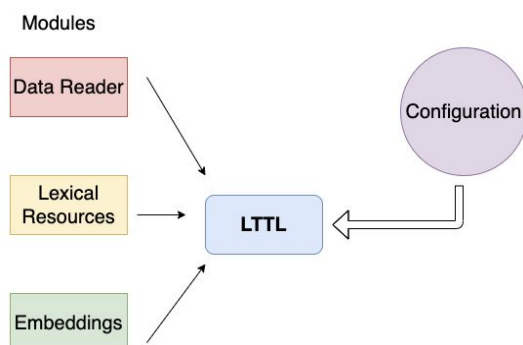


Figure 4: High-level overview of LTTL software architecture. The modules required by the LTTL learning framework are a data reader, lexical resources and embeddings; they are easily selected and parameterized using a configuration file.

The role of the *Lexical Resources* module is to integrate existing LLOD resources with the LTTL framework and allows for their pre-processing (deduplication, filtering) and composition. This module also provides for conversion scripts in order to render different data formats, so that, for experimentation, we can also integrate resources not (yet) available as LLOD.

The *Embeddings* module enables the use of different types of pre-trained monolingual embeddings with LTTL. Notably, it accepts embedding configurations with different dimensions for the source- and target-language embedding spaces. This allows us to flexibly experiment with combinations of pre-trained embeddings.

For handling and generating data sets, we created a *Data Reader* module that automatically interacts with Semalytix-internal corpora stored in a document database and prepares source- and target-language data sets. These datasets can be created for any available chosen class label and are saved in a JSON format to allow for easy creation and processing of a dataset for both languages. The creation of particular data sets based on existing corpora is entirely configuration-driven. The resulting data can directly be used with LTTL which applies all additional processing for training and evaluating a model (such as importing and autonomously splitting it into train-, test-, and development sets). An important ability of the reader is its compatibility with different language encodings that are not limited to Latin characters, which is especially important for transfer learning with other than Western-European languages. These abstractions, based on custom-created data sets in combination with a simple configuration step where concepts and tasks for LTTL are specified, enable language- and task-independent transfer learning. Any corpus for any language or task can be ingested via the same JSON format, allowing us to choose any desired language-task combination and customise it easily for individual experiments.

## 4 Language Resources

The LTTL framework requires two monolingual word embeddings spaces for the source and target language respectively, a bilingual lexicon with language pairs from source to target language, an annotated corpus for the source language labeled according to the task and an evaluation dataset in the target language. In this chapter, we describe the lexical resources and the procedures used to obtain them.

Up to this point, we have tried the LTTL framework in both sentiment and concept transfer tasks with European languages only. In our current experiments, English was used as the source language, and Spanish and French as target languages. In comparison, open-domain resources in English were easier to obtain due to English being a resource-rich language. On account of the biomedical or pharmaceutical nature of our use cases, we have also used domain-specific resources in this framework. Expectedly, pharma-domain resources were more difficult to find, especially in the target languages, and in some cases they were simply not available. Furthermore, the availability of bilingual lexica for the different language pairs varies greatly. In cases of appropriate resources not being available, we generated our own lexica via triangulation with one or more pivot languages. The procedure used to generate these is described in section 4.3.

### 4.1 Monolingual Word Embeddings

In the LTTL framework, the pre-trained monolingual embeddings are used for two purposes. The first is to obtain sentence representations in the source and target languages and the second is to represent translation pairs in a shared projection embedding space. In our experiments, we used pre-trained Word2Vec embeddings that were



publicly available and trained on large corpora. We distinguish between open-domain and domain-specific word embeddings. All embeddings used in our experiments are described in Table 1 below:

Embeddings	Language	Type	Vocabulary Size	Vector Dimensions
google-news	English	open-domain	55,627	300
PMC	English	domain-specific	2,515,686	200
Sg-300-es	Spanish	open-domain	834,213	300
scielo-wiki	Spanish	domain-specific	324,452	300
fr-wiki	French	open-domain	2,500,733	300

Table 1: Overview of monolingual embeddings used in experiments with the LTTL framework.

The vocabulary of open-domain embeddings contains words which are not particular to a specific domain. Because they can be used in multiple tasks, they are more easily available than domain-specific embeddings, especially for the English language. These resources are trained in open-domain texts such as Google News or Wikipedia. Finding embeddings in other languages was more difficult, but still feasible. The vocabulary of domain-specific embeddings contains words from the biomedical domain, such as diseases, medical specialties, drug names, symptoms, among others. The corpora used to train these embeddings mostly comprise biomedical articles and abstracts from domain-specific repositories such as PubMed, PubMedCentral (PMC) or Scielo. These embeddings were difficult to find in both source and target languages. Depending on the target language they may not be readily available.

## 4.2 Bilingual Translation Dictionaries

Bilingual translation dictionaries have an important role in LTTL. They are not only used for mapping words into a shared bilingual embedding space, but also to optimize the projected embeddings. During training, this happens by minimizing the semantic distance in the shared bilingual space between embedding vectors corresponding to translation pairs. Two types of lexica were used: single-source lexica and extended lexica. Single-source lexica provide translations from a single data source and extended lexica incorporate translations from two or more lexical resources. The lexica presented here vary in terms of vocabulary size, the type of knowledge they provide, origin, and purpose. At this point it should be noted that, as a consequence of the framework's architecture, only single word entries on both source and target languages are valid translation pairs that can be used productively in the tasks described above. Improvements to include multiword expressions and more linguistically informed sentence representations are to be researched in further work.

### 4.2.1. Single-source Lexica

Similarly to the monolingual embeddings, single-source lexica can be classified according to the type of knowledge they provide. In our experiments, we selected lexica according to the criteria of domain- and task specificity. Accordingly, broad- coverage open-domain lexica, a sentiment lexicon and a pharma lexicon were used as described in the following:

- Apertium lexica were used in various experiments. These are very comprehensive open-domain, broad-coverage lexica. They are available in multiple language pairs and include entries from different domains, thus making this the most linguistically rich and sophisticated resource out of the simple resources.

Originally, this resource was generated from an open-source machine translation platform with the same name. The available bilingual lexica were then converted into RDF and published as linked data by participants of this project.

- The sentiment lexicon was used on a sentiment transfer task for the languages English and Spanish, which is reported in section 5.1.. This is an open-domain, task specific resource originally provided as a monolingual resource by Hu and Liu. The monolingual variant English sentiment lexicon contains positive and negative opinion words and, in total, 5,749 entries . It was translated into Spanish by Barnes et al. using Google Translate to produce a bilingual EN-ES translation dictionary. Note that we do not make use of the polarity information provided for each entry.
- The pharma lexicon is a domain specific lexical resource that contains entries from the biomedical domain for the language pair English-Spanish only. This lexicon is the smallest and has a total of 2,687 entries which were extracted from bilingual entity lexicalizations from the Semalytix Knowledge Graph – an in-house repository of pharma-specific knowledge. Entity types such as diseases and symptoms, medical professions, drug products and agents, drug manufacturers, and therapy areas, among others are entries in this dictionary.

These lexica were pre-processed prior to being used in LTTL. Pre-processing comprises three steps:

- Deduplication: Duplicate entries are removed from the bilingual dictionary.
- Disambiguation: In case of translation ambiguities, i.e., more than one translation for a word, the translation candidate that occurs most frequently in the target language corpus (dependent on the task) is selected. This is done because there may be concurring translations for words inside the corpus. In this step, only those entries that are both present in the corpus and in the lexicon are disambiguated. Non-matches, i.e., translations that are not present in the target language corpus, are not discarded from the bilingual dictionary at this step.
- Filtering: In addition to the disambiguation step, all entries with translations in the bilingual dictionary that do not occur in the target language corpus are removed.

During our experiments we tried all these variants of preprocessing. The specific lexica used in the experiments are reported in sections 5.1. and 5.2. accordingly.

## 4.2.2. Extended Lexica

Lexicon extensions have as a starting point Apertium, here referred to as the base lexicon. Then, other lexical resources are added to the base lexicon. The goal of this extension procedure is to semantically enrich the shared bilingual embedding space by exploiting potential complementarities in lexical content from the other sources, since these enrich the base lexicon with other types of knowledge, such as domain or task-specific knowledge from the pharma lexicon and sentiment lexicon, respectively. For this purpose, individual source lexicons are composed successively in a given order with the base lexicon being the starting point. This is done by adding novel entries or overwriting existing ones, in case there are conflicting translations in the source lexicons. After the extension procedure, the novel lexical resource is processed according to the processing procedure described in 4.2.1.

As illustrated in Figure 5 below, we distinguish between three types of extensions: domain extensions, task extensions and full extensions. In the domain extension, domain vocabulary is added to the base lexicon. This means that entries with terms from the pharma domain are added to the base lexicon and in case of translation ambiguities, the translations from the pharmaceutical lexicon are given preference over the translations from the base lexicon. In the task extension, vocabulary related to the task (e.g., sentiment analysis or concept detection), is added to the base lexicon. Those entries from the sentiment lexicon are added to the base lexicon, and in case of translation ambiguities, the entries from the latter are preferred. The full extension adds both domain- and task specific vocabulary to the base

lexicon and in cases of translation ambiguities, the entries from lexical resources other than the base lexicon are preferred. Entries from the pharma lexicon are considered first, followed by entries from the sentiment lexicon.

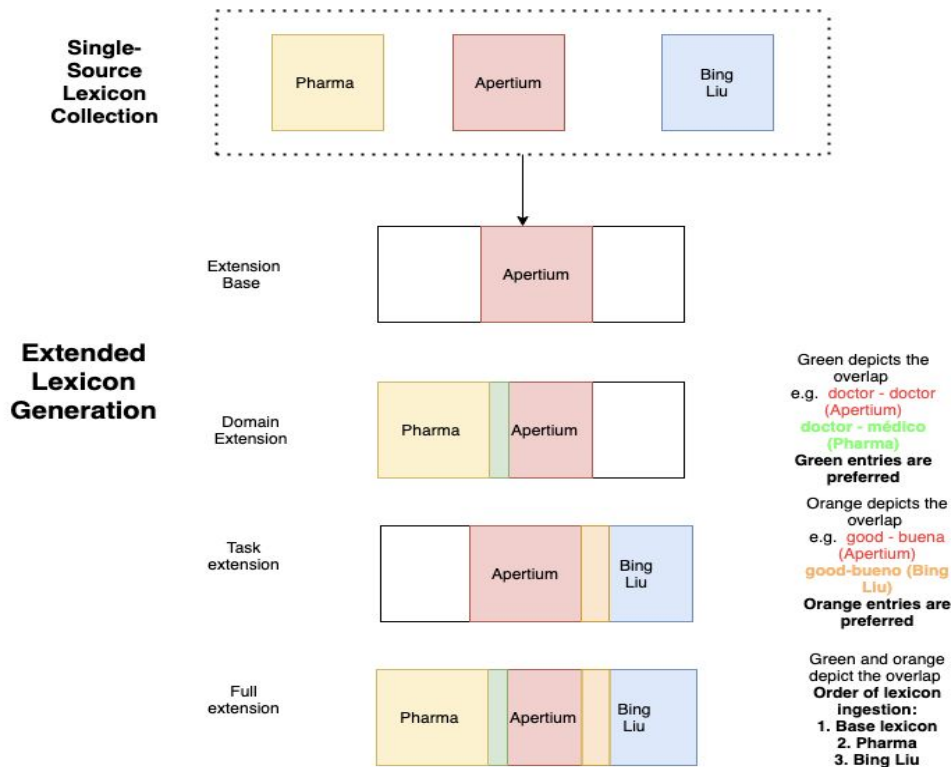


Figure 5: Extended lexicon generation procedure

This procedure enriches the base lexicon in different ways and this has implications for the configurations where these lexica are used and subsequently for the results. Furthermore, the further processing of the extended lexicon by deduplication, disambiguation and filtering (as described in section 4.2.1) play an important role in the configurations used in the different experiments. The different configurations used and their results are reported in section 5.

## 4.3 Cross-lingual Lexicon Induction

In some cases, bilingual lexica of interest for a given task or domain may not be available for all language pairs. One of the solutions for this problem is cross-lingual lexicon induction. This approach consists of leveraging two or more readily available lexical resources in either the source or target language and using a pivot language, i.e., a language which has correspondences to both languages, as a means to create a mapping between the two languages of interest.

More specifically, we used cross-lingual lexicon induction to bootstrap a bilingual dictionary for the language pair English-French using Apertium lexica either with one or multiple pivot languages via triangulation, as described in Figure 6 below for the example case of using Spanish as pivot.

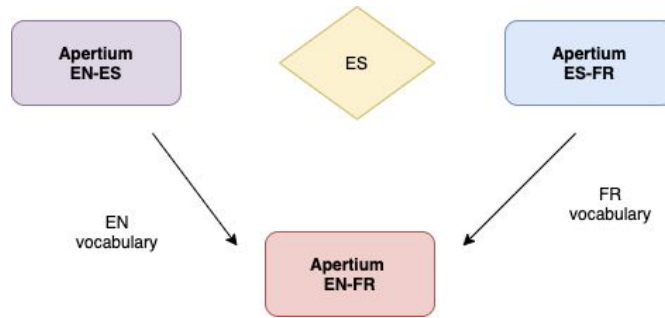


Figure 6: Lexicon induction procedure via a single pivot.

This procedure leverages corresponding entries between the pivot and the source and target language, respectively: For each entry that links a source language term  $t_s$  to its translation  $t_p$  in the pivot language, if there is an entry linking  $t_p$  to a target language term  $t_t$ , a translation from  $t_s$  to  $t_t$  can be inferred and stored in a newly created source-target lexicon. Subsequently, all duplicate entries are removed from the resulting lexicon.

Analogously, the same approach was used to create a bilingual dictionary via multiple pivots, as illustrated in Figure 7 below:

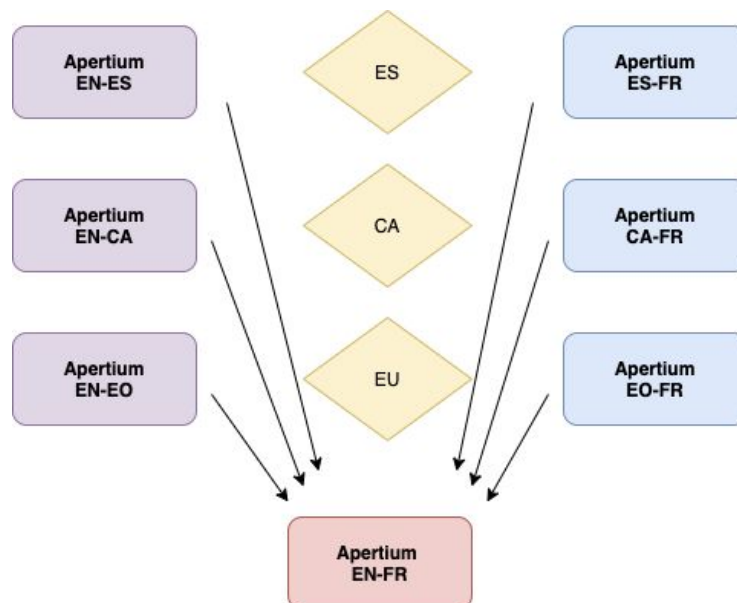


Figure 7: Lexicon induction procedure via multiple pivots.

Using multiple pivots results in lexically richer resources due to a higher number of entries and different types of entries provenient from the different sources. It also makes it possible to use the framework in lower resource languages, considering that Apertium lexica are available for many language pairs.

In addition to the vanilla cross-lingual induction approach, we added a pre-processing step, in which only entries that have the same part of speech (PoS) in both source and target language are included in the new bootstrapped lexicon. This procedure relies on the PoS information that is integrated into Apertium 2.0 via mapping lexical entries to the LexInfo ontology.

Table 2 below shows an overview of bootstrapped EN-FR lexica with either one or more pivot languages:

Lexicon	Number of entries
Apertium EN-FR with pivot ES	15,785
Apertium EN-FR with pivots ES and CA	60,447
Apertium EN-FR with pivots ES, CA and EO	94,195
Apertium EN-FR with pivots ES, CA and EO and PoS-preprocessing	70,012
Apertium EN-PT with pivots ES and CA	41,730

Table 2: Overview of lexica created with cross-lingual lexical induction and their corresponding number of entries.

As can be seen from the table, the more pivots are used, the more entries are part of the resulting lexicon. Also the PoS pre-processing provides an easy way to filter the entries so that they are more likely to be true correspondences. We postulate that words with the same PoS are more likely to be the correct translations because these words are used similarly in both languages. In return, we can see that this has a reducing effect on the lexicon. For comparison, we also induced a bilingual EN-PT lexicon using the same triangulation procedure (cf. last row in Table 2). This demonstrates, on the one hand, that it is easily feasible to scale our pipeline to other languages. On the other hand, comparing the statistics of the resulting lexicons suggests that the induced EN-FR lexicon may provide high lexical coverage to be used in LTTL. This hypothesis will be subjected to an empirical test in the experiments reported in the following.

The work reported in this section bears a strong connection to the TIAD campaign (“Translation Inference across Dictionaries”)<sup>1</sup> which is organized by UNIZAR, and the methods on lexical linking that are developed in WP3 (Task 3.2) of the Prêt-à-LLOD project. Until now, the procedures we apply in our work in order to infer translations for new language pairs are of baseline quality; in future work, we will seek collaboration with UNIZAR and other partners from WP3 in order to explore more sophisticated approaches to translation inference and capitalize on learnings from the TIAD task.

## 5 Experiments and Results

In this section we report the experiments and results for cross-lingual sentiment analysis and cross-lingual concept detection.

### 5.1 Cross-lingual Sentiment Analysis

The experiments we conducted for cross-lingual sentiment analysis using different lexical resources are described here. In these experiments, we investigated the impact of lexical resources on cross-lingual transfer of sentiment detection models. For this, we evaluated different configurations of lexical resources regarding their performance in terms of accuracy in the cross-lingual sentiment projection task.

---

<sup>1</sup> <https://tiad2020.unizar.es/>

### 5.1.1. Medical Experts' Transcripts

For our experiments, we used a corpus containing non-parallel samples of comparable English and Spanish medical expert transcripts. The transcripts contain summaries of conversations between pharma representatives and medical experts. In these conversations medical experts are asked to provide their opinions and assessments about certain aspects of medical treatments, for e.g., about the safety and effectiveness of drugs. The examples below denote positive and negative assessments of these aspects, respectively:

1. *DRUG can be safely used in elderly patients with renal failure.* (positive example, SAFETY aspect)
2. *No effect on glycaemic control when using DRUG as add-on.* (negative example, EFFECTIVENESS aspect)

In total, a collection of 21,400 English summaries was manually annotated with binary sentiment labels at the document level, resulting in 11,069 positive documents and 10,331 negative documents. These were used for training the cross-lingual transfer model in a cross-validation setting. Similarly, a set of 1,001 Spanish documents was annotated to create a test set for evaluation purposes only (559 positive documents, 442 negative documents).

### 5.1.2. Lexical resources and experiment configurations

We processed our simple and extended lexica according to the three step procedure introduced in section 4.2.1. Table 3 below shows the number of entries before and after processing:

	Source lexicons	Number of Entries (Original)	Number of Entries (Processed)
Single-source	Apertium	28,611	5,084
Single-source	BingLiu	5,749	1,362
Single-source	Pharma	2,687	277
Domain Extension	Apertium + Pharma	31,192	5,307
Task Extension	Apertium + BingLiu	34,254	5,799
Full Extension	Apertium + Pharma + BingLiu	36,941	6,018

Table 3: Overview of the lexical resources used in experiments with the LTTL framework with their corresponding number of entries before and after processing.

Based on previous experiment results obtained with these lexica, we paired them with the monolingual embeddings that provided the best performance. In previous experiments, we observed that Apertium benefits most from domain-specific embeddings in the target language (scielo-wiki), while the Pharma lexicon achieves the best results with open-domain embeddings in both the source (google-news) and target language (sg-300-es). The sentiment lexicon accomplishes together with domain-specific source embeddings (PMC) and open-domain embeddings in the target language (sg-300-es) the best results. In previous experiments, the combination of open-domain embeddings in

the source language and domain-specific embeddings in the target language yields the best results when paired with the different extensions.

### 5.1.3. Results

Table 4 shows the results of cross-lingual projection using the LTTL framework for each configuration of resources in terms of accuracy.

Lexicon	Monolingual Embeddings	Target language accuracy
Apertium	google; scielo_wiki	<b>0.768</b>
Pharma	google; sg300es	0.434
BingLiu	PMC; sg300es	0.711
Apertium + Pharma	google; scielo_wiki	0.763
Apertium + BingLiu	google; scielo_wiki	<b>0.773</b>
Apertium + Pharma + BingLiu	google; scielo_wiki	0.767

Table 4: Results of experiments with different lexicons and monolingual embedding combinations. For evaluation target language accuracy was measured.

If we compare the simple lexica, we can observe that the best accuracy in the target language occurs when Apertium is used (Acc = 0.768). This result supports the status of Apertium being a linguistically rich, general-purpose source of bilingual lexical knowledge. Despite the corpus being highly pharma-specific, the relative individual performance of the pharma and BingLiu lexica suggest that task-specific information is more important than technical domain knowledge. Furthermore we can observe the same complementarity of resources with the new processed lexica with respect to the monolingual embeddings used.

We also see that lexical extensions based on the Apertium lexicon can be effective and provide slightly better results than when using Apertium alone (Acc=0.773) due to their richer bilingual lexical representations. This seems to be due the complementarity among original lexica, since general-purpose knowledge from Apertium paired with task-specific knowledge from BingLiu achieve the best performance overall. As analyzed in more detail in Hartung et al. (2020), the best LTTL configuration obtained from these experiments outperforms a purely translation-based transfer approach by a wide margin (approx. 24 points in target language accuracy), and also comes close to the performance of a dedicated source language classification model (with a performance gap of only 5 points in source language accuracy).

## 5.2 Cross-lingual Concept Detection

Apart from sentiment analysis, we also conducted experiments for the task of concept detection across languages. We define this task as a text classification problem based on a variety of pharma-related output labels, e.g. the safety or effectiveness of a drug product, or health-related quality-of-life variables as reported by patients, such as their ability to perform their work, for instance. We approach the problem of predicting these concepts via a multitude of individual

binary classifiers. In the experiments reported in the following, LTTL is challenged to project such classifiers across languages.

### 5.2.1 Available monolingual model and generation of test set labels

In order to generate labeled data in the target language of interest for validation purposes, we rely on a heuristic label propagation procedure based on a monolingual classification model that is available in the Semalytix technology stack for processing English documents. It is a custom rule-based pattern matching engine that is largely based on manual rule engineering. There are two matching options for those rules: the literal matching of specific sequences of tokens in the text, or constraint-based matching that allows for additional complexity based on, e.g., regular expressions to capture morphological variation, part-of-speech tagging, dependency syntax or knowledge graph type constraints.

In order to make use of this monolingual model for texts that are not written in English, foreign language texts are algorithmically translated into English. Thus, the concept detector model can be run on the translated texts in the same way as on originally English ones. The resulting concept labels are then propagated back to the original documents. An illustration of this process is depicted in Figure 8 below.

Following this label propagation approach, we were able to generate sufficiently large testing samples for a great variety of concept detection problems, in the interest of subjecting LTTL to an extensive evaluation across individual tasks. It needs to be emphasized, though, that the resulting target-language labels were not manually checked for correctness. Hence, even though the underlying rule-based classifiers available for English are optimized for precision, the test collection resulting from this procedure must be considered a silver standard.

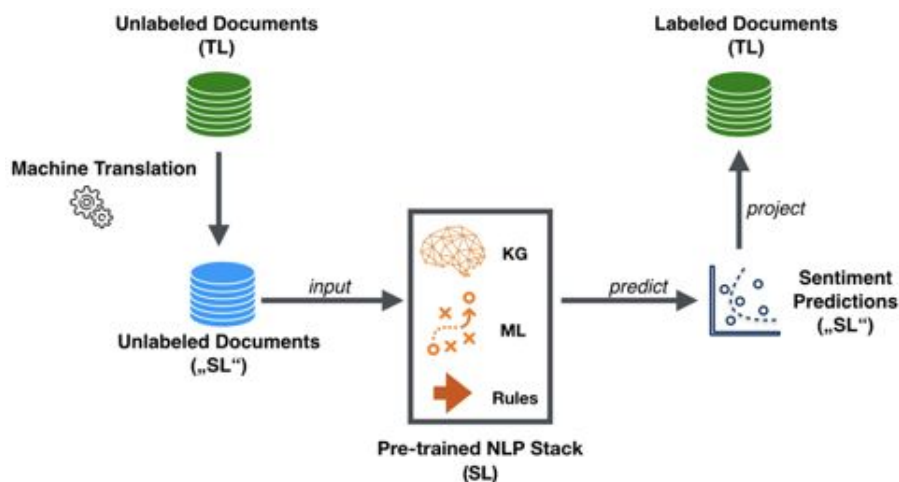


Figure 8: Label propagation from labeled source language to unlabeled target language documents.

### 5.2.2 Baseline Models

As comparison to our LTTL model, we also generate two baseline models:

- Baseline 1: Directly translate the rules used in the monolingual classifier into the target language, where they can subsequently be used as rule-based extractors.
- Baseline 2: Apply the monolingual classifier to English texts and, translate the resulting extractions, and subsequently use them as extractors in the target language.



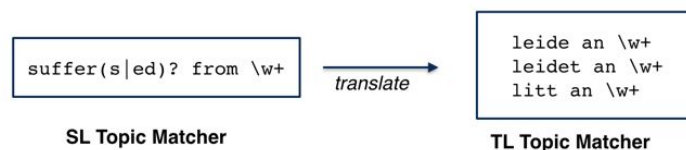


Figure 9: Illustration of pattern translation procedure of Baseline 1.

As illustrated in Figure 9, our approach for Baseline 1 (BL1) is to extract all patterns for each required concept in the source language (SL) and to then translate those to the target language (TL) using the DeepL translation API. After this step the resulting patterns are run on the target language silver standard test set. If a match is found, the given document is classified as a positive instance of the respective concept, otherwise a negative one.

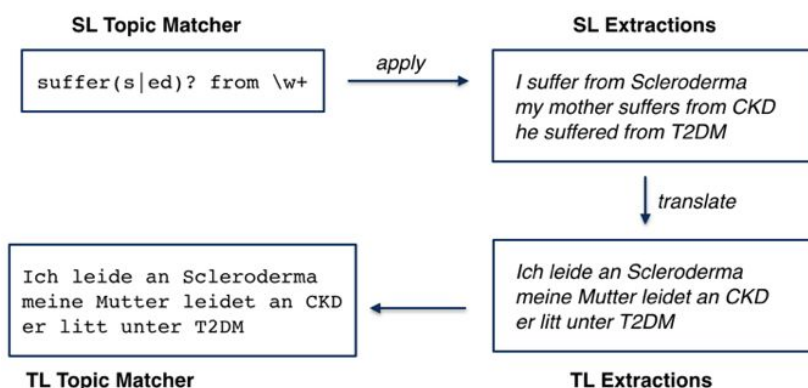


Figure 10: Illustration of pattern translation procedure of Baseline 2.

Baseline 2 (BL2) is following a slightly different approach (illustrated in Figure 10). First the original English patterns for each required concept are executed on the English training data set. Then, all English phrases that match those patterns are extracted and translated into the target language using the same translation service. Subsequently, those extractions (which in comparison to Baseline 1 do not usually contain any regular expressions or other formal constraints) are then used as target language extraction rules and run on the target language silver standard test set, analogously to Baseline 1.

### 5.2.3 Experiment 1: Medical Experts' Transcripts

The goal of this task is cross-lingual text classification on categorial labels from medical experts' transcripts data for the language pair English-Spanish. The corpus used in this experiment for cross-lingual concept detection consists of the same non-parallel sample of comparable English and Spanish conversations transcripts between pharma representatives and medical experts as for the cross-lingual sentiment analysis task (cf. section 5.1).

In this case it is particularly important that the medical experts in the data state their opinions and assessments about particular aspects related to the health outcomes of medical treatments. This results in the categorial labels from HCP insights concepts that our concept detection model aims to predict. For the same examples as in section 5.1 above, the labels to predict are SAFETY and EFFECTIVENESS:

- (1) DRUG can be safely used in elderly patients with renal failure. → SAFETY  
 (2) No effect on glycaemic control when using DRUG as add-on. → EFFECTIVENESS

An important characteristic of this data is that due to its nature as expert transcripts, it is rather structured and homogenous, especially compared to the social media data to be analysed in the next task (cf. 5.2.4), which allows for much more variation in the data samples.

To obtain the concept labels for the training, development and test sets, we ran the English-based Semalytix-internal concept detector for the two respective concepts on them. The labels for the Spanish data sets are propagated from their English translation, following the procedure outlined in section 5.2.1.

For the lexica and embeddings the same ones as in the cross-lingual sentiment analysis task were used as well, but in different configurations. Table 5 below gives an overview of the applied processed lexica and their number of entries. For more information on the different lexica versions see section 4.2 on the bilingual translation dictionaries.

Filter method	Lexicon combination	No. of entries
Deduplicated	Apertium	28,505
	Apertium+Pharma	30,708
Disambiguated	Apertium	25,355
	Apertium+Pharma	27,851
Filtered	Apertium	5,084
	Apertium+Pharma	5,361

Table 5: Number of entries for each EN-ES lexicon used in the experiments.

Table 6 below presents the results achieved by LTTL for the EN-ES language pair (listing only the best-performing combinations of monolingual word embeddings, and all configurations of bilingual lexicons as introduced in Table 5), as well as the two baselines described in section 5.2.2 above. Results are reported in terms of precision, recall and F1-measure for the positive class.

Overall, these results highlight an interesting difference across the two concepts investigated: While the translated rule-based baselines clearly perform best on the SAFETY concept, LTTL has an advantage on EFFECTIVENESS (with both baselines being strong competitors in terms of recall, though). On the latter concept, classification performance is lower in general across all approaches, which suggests a higher complexity of the task, and may also explain the relative superiority of the rule-based approach on SAFETY. In fact, qualitative analysis reveals EFFECTIVENESS as the more heterogeneous topic in terms of a greater linguistic variability. In addition, the volume of training data available seems to have a strong impact as well: It is approximately four times higher for EFFECTIVENESS than SAFETY, which may undermine the relative performance of LTTL on SAFETY even further.

In regard to the impact of the monolingual word embeddings, it is noticeable that the Spanish domain-specific *scielo\_wiki* embeddings do particularly well, while for English the general *google* ones achieve among the best results, with some of them yielded by the *PMC* domain-specific embeddings. Regarding the lexica, there is no clear winning configuration. The general single-source *Apertium* lexicon outperforms the combinations for the concept of EFFECTIVENESS, whereas it stays behind the domain extension of *Apertium+Pharma* for SAFETY. Considering lexica sampling methods no superior one can be detected either, as the table shows that on the one hand *disambiguated* obtains the best results for SAFETY (however, only in combination with *Apertium+Pharma*, while for just *Apertium* it is outperformed by *filtered*). On the other hand the simple *deduplicated* version of *Apertium* on its own, as well as *Apertium+Pharma* does best.

Lexicon		Embeddings	Precision	Recall	F1
<b>Safety</b>					
Apertium	Deduplicated	google + scielo_wiki	0.692	0.270	0.388
	Disambiguated	google + scielo_wiki	0.662	0.295	0.408
	Filtered	PMC + scielo_wiki	0.693	0.510	0.588
Apertium+Pharma	Deduplicated	google + scielo_wiki	0.686	0.285	0.402
	Disambiguated	PMC + scielo_wiki	0.787	0.500	0.612
	Filtered	PMC + scielo_wiki	0.695	0.525	0.598
Baseline 1			0.830	<b>0.760</b>	<b>0.793</b>
Baseline 2			<b>0.982</b>	0.560	0.713
<b>Effectiveness</b>					
Apertium	Deduplicated	google + scielo_wiki	<b>0.590</b>	0.655	<b>0.620</b>
	Disambiguated	google + scielo_wiki	0.584	0.605	0.594
	Filtered	google + scielo_wiki	0.588	0.600	0.594
Apertium+Pharma	Deduplicated	google + scielo_wiki	0.588	0.650	0.617
	Disambiguated	google + scielo_wiki	0.572	0.610	0.590
	Filtered	google + scielo_wiki	<b>0.590</b>	0.585	0.587
Baseline 1			0.481	0.585	0.528
Baseline 2			0.501	<b>0.680</b>	0.577

Table 6: Results for representative configurations of lexica and embeddings for EN-ES concept transfer in comparison to both baselines. Results are reported in terms of precision, recall and f1-measure for the positive class.

## 5.2.4 Experiment 2: Social Media Data Experiments

Taking into consideration the previous experiments and their results, we conducted another series of experiments which was designed in order to answer the following questions:

- Is it possible to apply LTTL across a variety of tasks, relying on little configuration efforts (cf. section 3.2), but under otherwise identical conditions?
- Can LTTL be effectively used on a language pair for which no bilingual lexicon is readily available, but needs to be induced beforehand?
- Can LTTL be effectively combined with rule-based approaches in order to increase classification performance?

For this set of experiments the setup of the cross-lingual concept detection task stays the same, but we now perform text classification on categorical labels from a WHO-based quality of life (QoL) taxonomy for the language pair

English-French. Figure 11 below gives an overview of the areas the QoL taxonomy comprises and the specific topics included in them. We conducted experiments on the 21 of them that are in use at Semalytix and thus have an available pattern detection model. *Self-esteem* and *Freedom, physical safety and security* were therefore excluded and *Religion, Spirituality* and *Personal beliefs* were grouped into one single concept.

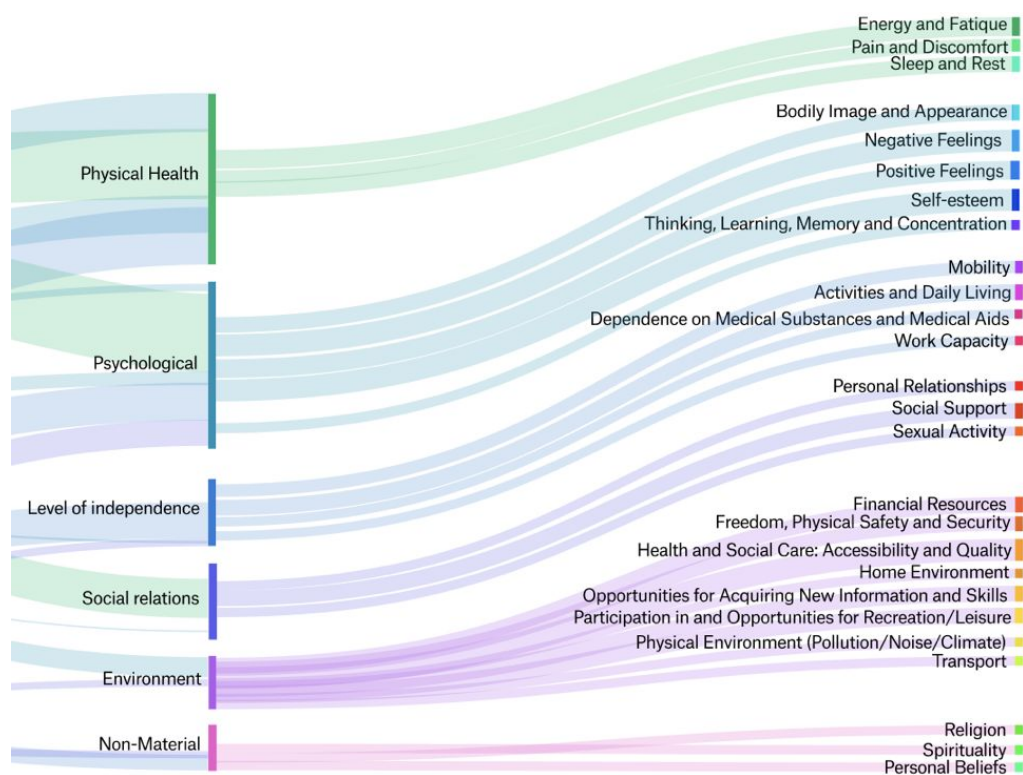


Figure 11: Overview of QoL areas and contained topics.

As data we used a social media corpus, made up of posts from several medical online fora. Compared to the medical transcripts used in the experiments in section 5.2.3, this data contains extremely varied language as it is not authored by medical experts, but often by patients and their relatives. This also poses challenges for concept detection via the Semalytix-internal concept detector, as it may be difficult to find precise but widely applicable rules.

Labeled data sets for both languages were created following the steps outlined in Section 5.2.1, i.e., tagging them via the English concept detector model and, in the case of the target language, propagating the obtained labels to the original French texts. Table 7 below gives an overview of the number of documents obtained this way for each concept. Numbers vary according to the available pattern detection results coming from the Semalytix-internal pipeline. As some concepts appear less frequently in the social media data corpus, their training and test sets contain fewer entries. All source data sets are automatically split into a training, development and test set (80/10/10) for training by the LTTL module. The target language sets are used for evaluation and thus automatically divided into a development and test set (50/50).

Concept	No. of documents in English data set per pos/neg samples	No. of documents in French data set per pos/neg samples
Activities of daily living	931	405
Body image and appearance	1333	904
Dependence on medicines	125	31
Energy and Fatigue	3000	2549
Financial resources	2451	859
Health and social care	1825	623
Home environment	840	287
Mobility	1141	456
Negative feelings	1685	1620
Pain and discomfort	3000	3000
Participation in and opportunities for recreation and leisure	3000	2718
Personal relationships	3000	3000
Physical environment	955	423
Positive feelings	3000	2277
Religion, spirituality and personal beliefs	3000	202
Sexual activity	50	132
Sleep and rest	823	523
Social support	893	466
Thinking, learning, memory and concentration	430	186
Transport	896	278
Work capacity	707	184

Table 7: Overview of number of documents for the source and target language data sets for all concepts for each positive and negative samples.

As the source language (English) is the same as in previous experiments, we used the same embeddings for it. For French, as we could not find suitable embeddings related to the pharmaceutical, medical or health domain, we employed the open domain open-access *frwiki* embeddings (for more details regarding the monolingual word embeddings see section 4.1.) Given that Apertium does not include an English-French lexicon, we created one via the pivot language Spanish (as the English-Spanish lexicon had worked well before) using the procedure described in section 4.3 on cross-lingual lexicon induction. We then also applied all filter methods to this lexicon, resulting in 3 versions of it. As no suitable domain-specific lexicon was available to us for FR-EN at the time of the experiments, we could not conduct them for the combination of Apertium with such a lexicon in the way we did for the Spanish data.

Furthermore, following the same approach as in our other experiments, we also obtained results for both baselines 1 and 2 for the EN-FR concept detection task. They will be discussed below in context of the results stemming from the experiment configurations of the different lexica and embedding combinations.

Additionally, following a detailed analysis into the classification predictions made by our different models, we explored the setting of combining both the baseline and the LTTL models in a sequential way. This was done by first executing the baselines and then LTTL, and vice versa. For each scenario, we ran the first model (Baseline 1 / 2 or LTTL) and then extracted all data (from both the positive and the negative samples) that had not been matched by this model.

The unmatched data sample was then given to the other model as input data. This results in 4 different configurations: BL1+LTTL, BL2+LTTL, LTTL+BL1 and LTTL+BL2. After obtaining results for the second model, evaluation metrics were calculated for the entire data set. We report these below in comparison to those of the single models for LTTL+BL, as the latter sequence outperformed BL+LTTL in nearly all cases.

Tables 8 and 9 below show the results of our experiments in terms of F1 measure, precision and recall for the positive class. The LTTL configuration employed the *google* embeddings for English, the *frwiki* ones for French and the *Apertium disambiguated* version of the EN-FR lexicon that was developed via a pivot language. This configuration was selected, as it yielded the overall best results in a first run for a subset of concepts on which we tested different lexicon and embedding configurations.

	Baselines						LTTL		
	BL 1			BL 2					
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Activities of daily living	0.562	0.425	<b>0.484</b>	0.569	0.482	<b>0.521</b>	0.718	0.616	<b>0.663</b>
Body image and appearance	0.904	0.146	<b>0.251</b>	0.937	0.381	<b>0.541</b>	0.829	0.527	<b>0.644</b>
Dependence on medicines	0.45	0.29	<b>0.353</b>	0.485	0.516	<b>0.5</b>	0.632	0.75	<b>0.686</b>
Energy and Fatigue	0.455	0.781	<b>0.575</b>	0.853	0.633	<b>0.727</b>	0.638	0.635	<b>0.636</b>
Financial resources	0.901	0.201	<b>0.329</b>	0.932	0.286	<b>0.438</b>	0.534	0.993	<b>0.695</b>
Health and social care	0.981	0.164	<b>0.281</b>	0.498	0.738	<b>0.595</b>	0.698	0.593	<b>0.641</b>
Home environment	0.519	0.801	<b>0.63</b>	0.953	0.143	<b>0.248</b>	0.596	0.646	<b>0.62</b>
Mobility	0.822	0.305	<b>0.445</b>	0.961	0.268	<b>0.419</b>	0.626	0.794	<b>0.7</b>
Negative feelings	0.955	0.106	<b>0.19</b>	0.955	0.157	<b>0.269</b>	0.629	0.642	<b>0.635</b>
Pain and discomfort	0.794	0.333	<b>0.469</b>	0.783	0.357	<b>0.49</b>	0.713	0.703	<b>0.708</b>
Participation in and opportunities for recreation and leisure	0.528	0.266	<b>0.353</b>	0.672	0.286	<b>0.402</b>	0.67	0.65	<b>0.66</b>
Personal relationships	0.503	0.601	<b>0.548</b>	0.669	0.588	<b>0.626</b>	0.666	0.679	<b>0.672</b>
Physical environment	0.498	0.56	<b>0.527</b>	0.983	0.404	<b>0.573</b>	0.722	0.712	<b>0.717</b>
Positive feelings	0.472	0.817	<b>0.598</b>	0.553	0.316	<b>0.402</b>	0.789	0.234	<b>0.36</b>
Religion, spirituality and personal beliefs	0.52	0.787	<b>0.626</b>	1	0.446	<b>0.616</b>	0.927	0.376	<b>0.535</b>
Sexual activity	0.976	0.614	<b>0.753</b>	0.975	0.598	<b>0.742</b>	0.464	0.394	<b>0.426</b>
Sleep and rest	0.941	0.182	<b>0.304</b>	0.927	0.507	<b>0.655</b>	0.642	0.473	<b>0.545</b>
Social support	0.929	0.028	<b>0.054</b>	0.75	0.013	<b>0.025</b>	0.678	0.579	<b>0.625</b>
Thinking, learning, memory and concentration	0.542	0.629	<b>0.582</b>	0.987	0.403	<b>0.573</b>	0.543	0.473	<b>0.506</b>
Transport	0.485	0.795	<b>0.602</b>	0.555	0.543	<b>0.549</b>	0.703	0.647	<b>0.674</b>
Work capacity	1	0.071	<b>0.132</b>	1	0.158	<b>0.272</b>	0.873	0.598	<b>0.71</b>

Table 8: Results for EN-FR concept transfer with LTTL in comparison to both baselines. Results are reported in terms of precision, recall and f1-measure for the positive class.

Table 8 depicts results for both baselines models in comparison to LTTL. While LTTL surpasses both baselines for a large number of concepts, in some cases (6 out of 21 concepts), it is actually outperformed by one of them (with BL1 achieving better results than BL2, apart from the case of *Sleep and rest*).

An observation that holds true for a majority of the concepts ( approx. 62%) is that baseline 2 regularly exceeds baseline 1 and even when it is surpassed by BL1, it is mostly not by a large margin (except for *Home environment*). Being designed as precision-oriented extraction rules for English documents, most of the baselines still favour precision after being transferred to French: For roughly 75% of the concepts at least one of the baselines shows a significantly better precision than LTTL. However, apart from a small number of cases, LTTL benefits from a much higher recall, which results in a better overall performance of this model. An explanation for this could be that the patterns employed in the concept detection model that the baselines are based on, loses coverage due to translation, but does usually not match more noise.

	Combinations					
	LTTL+BL1			LTTL+BL2		
	Precision	Recall	F1	Precision	Recall	F1
Activities of daily living	0.627	0.827	<b>0.714</b>	0.609	0.825	<b>0.701</b>
Body image and appearance	0.822	0.587	<b>0.685</b>	0.848	0.732	<b>0.786</b>
Dependence on medicines	0.578	0.839	<b>0.684</b>	0.6	0.968	<b>0.741</b>
Energy and Fatigue	0.5	1	<b>0.667</b>	0.68	0.933	<b>0.787</b>
Financial resources	0.533	0.999	<b>0.695</b>	0.533	0.999	<b>0.695</b>
Health and social care	0.719	0.674	<b>0.696</b>	0.527	0.923	<b>0.671</b>
Home environment	0.531	0.951	<b>0.682</b>	0.632	0.753	<b>0.687</b>
Mobility	0.658	0.908	<b>0.763</b>	0.678	0.904	<b>0.774</b>
Negative feelings	0.643	0.693	<b>0.667</b>	0.647	0.709	<b>0.677</b>
Pain and discomfort	0.702	0.844	<b>0.766</b>	0.698	0.853	<b>0.768</b>
Participation in and opportunities for recreation and leisure	0.617	0.79	<b>0.693</b>	0.661	0.815	<b>0.73</b>
Personal relationships	0.548	0.939	<b>0.692</b>	0.627	0.932	<b>0.75</b>
Physical environment	0.558	0.887	<b>0.685</b>	0.755	0.844	<b>0.797</b>
Positive feelings	0.496	0.954	<b>0.652</b>	0.621	0.528	<b>0.57</b>
Religion, spirituality and personal beliefs	0.558	0.936	<b>0.699</b>	0.96	0.718	<b>0.822</b>
Sexual activity	0.697	0.871	<b>0.774</b>	0.695	0.864	<b>0.77</b>
Sleep and rest	0.699	0.612	<b>0.652</b>	0.737	0.793	<b>0.764</b>
Social support	0.69	0.635	<b>0.661</b>	0.688	0.633	<b>0.659</b>
Thinking, learning, memory and concentration	0.534	0.844	<b>0.654</b>	0.647	0.72	<b>0.682</b>
Transport	0.523	0.957	<b>0.676</b>	0.595	0.86	<b>0.703</b>
Work capacity	0.868	0.609	<b>0.716</b>	0.872	0.63	<b>0.732</b>

Table 9: Results for EN-FR concept transfer with the combined pipelines for LTTL+Baseline. Results are reported in terms of precision, recall and f1-measure for the positive class.

While we also conducted experiments for Baselines 1 and 2 in sequential combination with LTTL, in Table 9 we only report LTTL+Baselines 1 and 2 as the latter repeatedly outperformed the former combination. The presented results illustrate that there are common tendencies regarding which model obtains the overall best results: For the majority of concepts the performance of the LTTL+Baselines combination exceeds LTTL and is among the best configurations, apart from *Financial resources*, where both models perform equally well. Winning results are



obtained by the combined models of LTTL+Baseline 1 or 2 by a sometimes considerable margin compared to LTTL for 20 out of 21 cases.

Another observation that holds true for many of the concepts, is that Baseline 2 regularly outperforms Baseline 1 not only when employed by itself, but also when used in sequence with LTTL (LTTL+BL2). Even when it is left behind Baseline 1, it is inferior only by a small margin. The only exception to this (*Home environment*) nevertheless has LTTL+B2 yielding better results. This likely stems from the advantage that the recall generally increases in the combined models, as the baselines add to the LTTL results in a sequential way (the extended coverage here seems to outweigh the potentially added noise).

This shows that the LTTL model undoubtedly benefits from the sequential addition of the baseline models. Leading results are obtained by the sequential combination of LTTL+BL2 for nearly 80% of the concepts and by LTTL+BL1 in the other approximately 20%, except for the one case where no gain over LTTL alone has been achieved.

Summing up, our LTTL system has proven to be applicable to a variety of tasks and concepts with little configuration efforts, thanks to its modular and easily configurable architecture. Moreover, we showed that it can be employed even when a bilingual lexicon for a particular language pair is not readily available, as it can be induced via one or more pivot languages without resulting in poor performance. Furthermore, the LTTL model can effectively be combined sequentially with rule-based concepts detectors, resulting in a noticeable increase of classification performance.

## 6 Summary and Conclusions

In this report, we presented the LTTL framework as a generalized transfer architecture which can be flexibly used in order to induce bilingual task-specific embeddings as lexical representations for NLP models that are needed for multilingual text analytics. Being embedded into an LLOD exploitation pipeline based on Prêt-à-LLOD core technology, LTTL is flexibly applicable to different languages and for various tasks, which we successfully demonstrated for the two tasks of sentiment analysis and health-related concept detection, both of which are of high practical relevance for multilingual text analytics in the pharmaceutical domain. Thus, the experiments reported provide evidence for LLOD lexical resources serving as catalyzers of cross-lingual transfer approaches which facilitate the scalability of real-world industrial solutions to a large variety of languages that are of relevance for global markets. In addition, we consider the outcomes of this pilot as corroborating the strong potential of the Prêt-à-LLOD value chain as an enabler of rapid model transfer frameworks, which may also involve rapid development of exhaustive benchmarking frameworks based on all available LLOD resources satisfying certain criteria (and possibly their combinations).

The current achievements of Pilot IV as reported in this document are in line with the planned progress according to the specification outlined in the Pilot Specification Report (Deliverable D4.1), as can be seen from the status overview of the involved milestones presented in Table 10 in Appendix A. In future work, we plan to refine and extend the currently existing LTTL workflows with a particular emphasis on bilingual lexicon induction and extension (cf. Section 4.3), as well as to integrate them into a running Pharos demo platform for evaluation and demonstration purposes.

## References

Julian Barnes, Roman Klinger, Sabine Schulte im Walde (2018): Bilingual sentiment embeddings. Joint projection of sentiment across languages. In: Proceedings of ACL.



Jorge Gracia, Christian Fäth, Matthias Hartung, Max Ionov, Julia Bosque-Gil, Susana Verissimo, Christian Chiarcos, Matthias Orlikowski (2020): Leveraging Linguistic Linked Data for Cross-Lingual Model Transfer in the Pharmaceutical Domain. In: J. Z. Pan et al. (eds.) The Semantic Web – ISWC 2020. Lecture Notes in Computer Science, vol 12507. Springer, Cham. [https://doi.org/10.1007/978-3-030-62466-8\\_31](https://doi.org/10.1007/978-3-030-62466-8_31)

Matthias Hartung, Matthias Orlikowski, Susana Verissimo (2020): Evaluating the impact of bilingual lexical resources on cross-lingual sentiment projection in the pharmaceutical domain. Technical Report. <https://doi.org/10.5281/zenodo.3707940>

Anders Søgaard, Ivan Vulic, Sebastian Ruder, Manaal Faruqi (2019): Cross-lingual word embeddings. Morgan Claypool.

## Appendix A

Milestone ID	Goal	Due date (Project Month)	Status (effective M24)
MS-IV.1	Prototype implementation of language transfer for at least one type of component (e.g. supervised ML) and at least one pair of source and target language (English and e.g. Spanish)	M9	achieved
MS-IV.2	Individually implemented language transfer of all analytical components needed to populate a complete dashboard	M15	achieved
MS-IV.3	Discover and consume cross-lingual LLOD resources and services supporting manual language transfer of a complete dashboard	M18	achieved
MS-IV.4	Language transfer pipeline consuming LLOD resources for at least <i>one type of analytical component</i> ; Pilot Report Version 1	M24	achieved
MS-IV.5	Language transfer pipelines consuming and transforming LLOD resources for <i>all analytical components</i> needed to populate a complete dashboard	M32	in progress
MS-IV.6	Evaluations complete	M33	to do
MS-IV.7	Documentation complete	M35	to do
MS-IV.8	Pilot Report Version 2	M36	to do

Table 10: Milestones and timeline for Pilot IV from the Prêt-à-LLOD Pilot Specification Report