

CLASSiC

D2.5: Improved models of syntactic-semantic parsing, for English and French.

Lonneke van der Plas, James Henderson, Paola Merlo

Distribution: Public

CLASSiC

Computational Learning in Adaptive Systems for Spoken Conversation
216594 Deliverable 2.5

February 2011



Project funded by the European Community
under the Seventh Framework Programme for
Research and Technological Development



CogSys
Cognitive Systems



The deliverable identification sheet is to be found on the reverse of this page.

Project ref. no.	216594
Project acronym	CLASSiC
Project full title	Computational Learning in Adaptive Systems for Spoken Conversation
Instrument	STREP
Thematic Priority	Cognitive Systems, Interaction, and Robotics
Start date / duration	01 March 2008 / 36 Months

Security	Public
Contractual date of delivery	M36 = Feb 2011
Actual date of delivery	February 2011
Deliverable number	2.5
Deliverable title	D2.5: Improved models of syntactic-semantic parsing, for English and French.
Type	Prototype
Status & version	Final 1.0
Number of pages	13 (excluding front matter)
Contributing WP	2
WP/Task responsible	UNIGE
Other contributors	
Author(s)	Lonneke van der Plas, James Henderson, Paola Merlo
EC Project Officer	Philippe Gelin
Keywords	Syntactic-semantic parsing

The partners in CLASSiC are:

Heriot-Watt University	HWU
University of Cambridge	UCAM
University of Geneva	GENE
Ecole Supérieure d'Electricité	SUPELEC
France Telecom/ Orange Labs	FT
University of Edinburgh HCRC	EDIN

For copies of reports, updates on project activities and other CLASSiC-related information, contact:

The CLASSiC Project Co-ordinator:
Dr. Oliver Lemon
School of Mathematical and Computer Sciences (MACS)
Heriot-Watt University
Edinburgh
EH14 4AS
United Kingdom
O.Lemon@hw.ac.uk
Phone +44 (131) 451 3782 - Fax +44 (0)131 451 3327

Copies of reports and other material can also be accessed via the project's administration homepage, <http://www.classic-project.org>

©2011, The Individual Authors.

No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

Contents

Executive Summary	1
1 Overview	2
2 Domain-general syntactic-semantic parsing for several languages	2
2.1 The probability model	3
2.2 The learning architecture	3
2.3 Experiments	3
2.4 Extension to other languages	4
3 Building a French corpus annotated with semantics by porting annotation	4
3.1 Models of cross-lingual semantic transfer	5
3.2 Experiments	5
4 Adaptation of an English parser to goal-directed spoken data	7
4.1 Creating spoken data automatically	7
4.2 Experiments	8
5 Conclusions	9
6 Running the Prototype Syntactic-Semantic Parsers for English and French	10

Executive summary

This document describes the Prototype deliverable 2.5, due at month 36 of the CLASSiC project. The prototype is an improved syntactic-semantic parsing architecture, applied to both English and French.

This document presents an overview of the steps we took to improve the parsing framework, how it was applied to French using cross-lingual transfer of semantic annotations, and its adaptation to spoken data. All steps, adaptations and applications have been evaluated and this document shows the main results. We also provide information on how to run the prototype.

To preview, UNIGE developed a parsing architecture for joint syntactic-semantic dependency parsing, based on Incremental Sigmoid Belief Networks (ISBNs), a latent variable model for syntactic structure prediction. We participated in the CoNLL-2008 shared task on English syntactic and semantic parsing, where it yielded 79.1% macro-average F1 performance, for the joint task, 86.9% syntactic dependencies LAS and 71.0% semantic dependencies F1. The ability of ISBNs to induce their features automatically enables us to extend this architecture to syntactic-semantic parsing of other languages without hand-crafted feature-engineering, as confirmed by the results from the CoNLL-2009 shared task on syntactic and semantic parsing of seven languages [1]: The system was ranked third overall with a macro averaged F1 score of 82.14%, only 0.5% worse than the best system.

To be able to aid in French SLU systems developed in the framework of the CLASSiC project we needed to train a French syntactic-semantic parser. Encouraged by the good results attained when parsing several other languages from the same language family, such as Catalan and Spanish, as part of the CoNLL-2009 shared task, we used the same parsing architecture. Because there is currently no semantically annotated corpus for French and hand annotation is very time consuming, we developed a method for cross-lingual automatic transfer as explained in Deliverable 6.2. We improved the method in several ways and are now able to generate automatically annotated data that is very close in quality to manual annotations.

Lastly, we discuss our work on domain adaptation. We adapted a constituency parser based on the same parsing framework to the domain of spoken dialogue conversation, by hand-annotating a small sample of goal-directed spoken data. On the basis of regularities and distributions found in the sample we automatically generated a large amount of training data using a three-component model. We show that we are able to improve the performance of a statistical parser on goal-directed spoken data extracted from human-machine dialogues without degrading the performance on full sentences.

Some aspects of this work were published at ACL 2009 [2], at NAACL 2009 [3], at CoNLL 2008 [4] and 2009 [5] and at IJCAI 2009 [6], and the ACL Linguistic Annotation Workshop [3].

1 Overview

Task 2.5 investigates improved statistical models of semantic parsing which incorporate linguistically motivated analyses of syntactic and semantic structure. What in NLP is called Semantic Parsing is based on a task of Semantic Role Labelling (SRL). SRL consists in assigning abstract labels to participating individuals in a scenario-like frame expressed by a sentence. For example, given the sentence *She blames the government for her failure*, we want to label the noun phrases (*she*, *government*) as (JUDGE, EVALUEE) respectively, and disambiguate the prepositional phrase *for failure* as REASON. We also identify the verb *blame* as a verb of JUDGEMENT. Semantic role labelling has been shown to be useful in several user-interface tasks, such as question answering, and could provide an abstract representation to dialogue exchanges.

UNIGE has developed a domain-general syntactic-semantic parsing framework that uses a generative history-based latent variable model to predict the most likely derivation of a synchronous dependency parser for both syntactic and semantic dependencies. Results from the participation of UNIGE in the CoNLL-2008 shared task [4] indicate that synchronous parsing is an effective way of building joint models on separate structures. In [6], the model has been improved to better handle non-planarity in semantic dependency graphs. The probabilistic model is based on Incremental Sigmoid Belief Networks (ISBNs), a latent variable model that induces its features automatically. The generality of the ISBN design suggests that ISBN's latent feature induction extends well to estimating very complex probability models, with little need for feature engineering. Results from UNIGE's participation in the CoNLL-2009 shared task confirmed the flexibility of ISBN's latent feature induction with very good performance on many languages without the need for manual feature engineering. Moreover, the joint syntactic-semantic parsing framework, has proven beneficial for the task of automatic annotation generation for French, where it was able to correct automatic transfers of semantic annotation by joint learning of syntactic and semantic structures. A syntactic-semantic parser for French has been trained on these annotations, thereby eschewing the need of manually annotating a corpus with semantic structure. As a last proof of the flexibility of the parsing framework, we successfully adapted an English syntactic-semantic constituency parser based on the same parsing architecture to the domain of goal-directed spoken data.

We break our presentation of this prototype into three subtasks: development, training and evaluation of the statistical syntactic-semantic dependency parser for English and its extension to six other languages; automatic generation of a French corpus of semantic annotations for training a French statistical syntactic dependency parser and its evaluation; adaptation of an English statistical syntactic-semantic constituency parser to the domain of goal-directed spoken data extracted from human-machine dialogues.

2 Domain-general syntactic-semantic parsing for several languages

Successes in syntactic tasks, such as statistical parsing and tagging, have recently paved the way to statistical learning techniques for levels of semantic representation, such as jointly learning the syntactic structure of the sentence and the propositional argument-structure of its main predicates [7, 8]. In this vein, the CoNLL 2008 shared task sets the challenge of learning jointly both syntactic dependencies, extracted from the Penn Treebank [9]) and semantic dependencies (extracted both from PropBank [10] and NomBank [11] under a unified representation.

UNIGE proposes a solution that uses a generative history-based model to predict the most likely derivation of a synchronous dependency parser for both syntactic and semantic dependencies [4].

2.1 The probability model

By solving the problem with synchronous parsing, a probabilistic model is learnt which maximises the joint probability of the syntactic and semantic dependencies and thereby guarantees that the output structure is globally coherent, while at the same time building the two structures separately.

We devise separate derivations $D_d^1, \dots, D_d^{m_d}$ and $D_s^1, \dots, D_s^{m_s}$ for the syntactic and semantic dependency structures, respectively, and then divide each derivation into the chunks between shifting each word onto the stack, $c_d^t = D_d^{b_d^t}, \dots, D_d^{e_d^t}$ and $c_s^t = D_s^{b_s^t}, \dots, D_s^{e_s^t}$, where $D_d^{b_d^{t-1}} = D_s^{b_s^{t-1}} = \text{shift}_{t-1}$ and $D_d^{e_d^{t+1}} = D_s^{e_s^{t+1}} = \text{shift}_t$. The actions of the synchronous derivations consist of quadruples $C^t = (c_d^t, \text{switch}, c_s^t, \text{shift}_t)$, where *switch* means switching from syntactic to semantic mode. This gives us the following joint probability model, where n is the number of words in the input.

$$\begin{aligned} P(T_d, T_s) &= P(C^1, \dots, C^n) \\ &= \prod_t P(C^t | C^1, \dots, C^{t-1}) \end{aligned} \quad (1)$$

The probability of each synchronous derivation chunk C^t is the product of four factors, related to the syntactic level, the semantic level and the two synchronising steps.

$$\begin{aligned} P(C^t | C^1, \dots, C^{t-1}) &= \\ &P(c_d^t | C^1, \dots, C^{t-1}) \times \\ &P(\text{switch} | c_d^t, C^1, \dots, C^{t-1}) \times \\ &P(c_s^t | \text{switch}, c_d^t, C^1, \dots, C^{t-1}) \times \\ &P(\text{shift}_t | c_d^t, c_s^t, C^1, \dots, C^{t-1}) \end{aligned} \quad (2)$$

2.2 The learning architecture

The crucial intuition behind the treatment of both syntax and semantic in a single model is that these two levels of information are related but not identical. We propose a solution that uses a generative history-based model to predict the most likely derivation of a synchronous dependency parser for both syntactic and semantic dependencies. Our probabilistic model is based on Incremental Sigmoid Belief Networks (ISBNs), a recently proposed latent variable model for syntactic structure prediction, which has shown very good behaviour for both constituency [12] and dependency parsing [13]. The ability of ISBNs to induce their features automatically enables us to extend this architecture to learning a synchronous parse of syntax and semantics without modification of the main architecture.

2.3 Experiments

We train and evaluate the models described above (augmented with a swap action to deal with crossing dependencies as indicated in [6]) on data provided for the CoNLL-2008 shared task on joint learning of syntactic and semantic dependencies. The data is derived by merging a dependency transformation of the Penn Treebank with Propbank and Nombank [14].

In our experiments, we use the measures of performance used in the CoNLL-2008 shared task, typical of dependency parsing and semantic role labelling. Syntactic performance is measured by percentage of

MODEL	CoNLL MEASURES			CROSSING ARCS		
	Syntactic	Semantic	Macro	Semantics		
	Labelled Acc.	F ₁	F ₁	P	R	F ₁
Johansson and Nugues 2008	89.3	81.6	85.5	67.0	44.5	53.5
UNIGE <i>Swap</i> 2009	87.5	76.1	81.8	62.1	29.4	39.9
UNIGE CoNLL 2008	87.6	73.1	80.5	72.6	1.7	3.3

Table 1: Syntactic and semantic labelling results of two versions of our system compared to the best result in the CoNLL-2008 shared task.

correct labelled attachments (LAS in the tables) and semantic performance is indicated by the F-measure on precision and recall on semantic arcs (indicated as SRL measures in the tables). These two components are then averaged in a score called Macro F₁.

The results are shown in Table 1. The second row reports the official CoNLL 2008 shared task numbers [4] and the third row shows the performances when using *Swap* reported in [6]. We compare to the best performing model for English [15] (first row).

While the already competitive syntactic performance of [4] is not significantly degraded, we notice an improvement of 3% on the semantic graphs when using the swap operation. This score approaches those of the best systems. As the right-hand panel on crossing arcs indicates, this improvement is due to better recall on crossing arcs. Also, importantly, this model is one of the few that does joint learning, with the best results in that category. The best performing system learns the two representations separately, with a pipeline of state-of-the-art systems, and then reranks the joint representation in a final step [15].

2.4 Extension to other languages

UNIGE has participated again in the CoNLL shared task of 2009 on syntactic and semantic dependency parsing for seven languages, with the syntactic-semantic dependency parser described in [6]. Good performance on multiple languages with a single model is particularly important for the CLASSiC project because we need a semantic parser that will also work well for French.

In [5] we evaluated the ability of this method to generalise across several languages. The model is taken as it was developed for English, and applied directly to all seven languages. The only fine-tuning done was to evaluate whether to include one feature type which we had previously found did not help for English, but helped overall. No other feature engineering was done. The use of latent variables to induce features automatically from the data gives our method the adaptability necessary to perform well across all seven languages, and demonstrates the lack of language specificity in the models of [4] and [6].

This system was ranked third overall with a macro averaged F1 score of 82.14%, only 0.5% worse than the best system.

3 Building a French corpus annotated with semantics by porting annotation

Broad-coverage semantic annotations for training statistical parsers are available for a handful of languages, for example for those that are part of the CoNLL-2009 shared task described in the previous

section, but not for French. In the CLASSiC project we are working on English and French SLU. We need to develop a syntactic-semantic parser for French to be used in French SLU and we therefore need semantically annotated data. Manually annotating sentences is costly. One approach to addressing this problem is to develop methods that automatically generate annotated data by transferring annotations in parallel corpora from languages for which this information is available to languages for which these data are not available [16, 17, 18].

In this line of work we aim to generate high-quality broad-coverage semantic annotations using an automatic, knowledge-lean approach, with the goal of training a broad-coverage syntactic-semantic parser for French using the syntactic-semantic parsing architecture described in section 2. We believe that syntactic information can be used to correct and complement automatic semantic transfers. Since we explained the method for transfer in previous deliverables (2.3 and 6.2), we limit ourselves here to a short description and mainly focus on the improvements we made.

3.1 Models of cross-lingual semantic transfer

Following the Direct Correspondence Assumption for syntactic dependency trees by Hwa et al. [19], our method for transferring semantic annotation from English to French is based directly on automatically generated word-alignments in the following way:

Direct Semantic Transfer (DST) For any pair of sentences E and F that are translations of each other, we transfer the semantic relationship $R(x_E, y_E)$ to $R(x_F, y_F)$ if and only if there exists a word-alignment between x_E and x_F and between y_E and y_F , and we transfer the semantic property $P(x_E)$ to $P(x_F)$ if and only if there exists a word-alignment between x_E and x_F .

The relationships which we transfer are semantic role dependencies and the properties are predicate senses. The word-alignments are those output by word alignment methods developed for statistical machine translation [20]. We introduce one constraint to the direct semantic transfer. Because the semantic annotations in the target language are limited to verbal predicates, we only transfer predicates to words the syntactic parser has tagged as a verb.

Because the direct correspondence assumption is a strong hypothesis that is useful to trigger a process of role projection, but will not work correctly for several cases, we introduce a filter. We know from the annotation guidelines used to annotate the French gold sentences that all verbs, except modals and realisations of the verb *être*, should receive a predicate label. We define a filter that removes sentences with missing predicate labels, referred to as the target filter.

3.2 Experiments

To transfer semantic annotation from English to French, we used the Europarl corpus [21]. Datasets were tokenised and lowercased and only sentence pairs corresponding to a one-to-one sentence alignment with lengths ranging from one to 40 tokens on both French and English sides were considered. Furthermore, we select only those parallel sentences in Europarl for which we know they are direct translations from English to French, or from French to English. We also excluded announcements from the data. After this pre-processing step, we are left with roughly 280-thousand sentence pairs. We word-align the English sentences to the French sentences automatically using GIZA++ [20] based on intersective alignment. For

		Predicates						Arguments (given predicate)					
		Labelled			Unlabelled			Labelled			Unlabelled		
		Prec	Rec	F	Prec	Rec	F	Prec	Rec	F	Prec	Rec	F
1	Transfer (no filter)	50	30	38	94	57	71	58	45	51	70	55	62
2	Transfer (target filter)	50	44	47	95	85	90	60	46	52	72	56	63
3	Parsing (no filter)	68	26	37	98	37	54	78	56	65	88	63	74
4	Parsing (target filter)	56	46	51	96	80	87	69	49	58	83	58	68
5	Inter-annotator agreement	61	57	59	97	89	93	73	75	74	88	91	89

Table 2: Percent recall, precision, and F-measure for predicates and for arguments given the predicate, for two transfer models, the two parsing models and the manual annotation.

testing we used the hand-annotated data described in [22] and D6.2. We randomly split the 1000 sentences into test and development set containing 500 sentences each.

We evaluate our methods for automatic annotation generation after the transfer step and after parsing. The comparison of these two steps will tell us whether the joint syntactic-semantic parser is able to improve semantic annotations by learning from the syntactic annotations available.

Table 2 shows the results of automatically transferring semantic annotation from English to French sentences. The first set of columns of results reports labelling and identification of predicates and the second set of columns reports labelling and identification of arguments respectively for the predicates that are identified. The first row shows the results when applying the direct semantic transfer without filtering. The second row shows results when using the target filter. Results from evaluating the parser trained on automatic transfers are shown in rows three and four.

The most striking result is that the scores after parsing are higher than after transfer. The parser is able to outperform the quality of the semantic data on which it was trained. This is especially clear for argument labelling, where the F-measure for the filtered setting increases from 53% to 58% after parsing (51% versus 65% for the unfiltered data). We believe that this is due to the fact that we are working in a joint syntactic-semantic setting. The parser has access to semantic and syntactic annotations (each from a different source) and uses the two types of information to build a mutually-informed probabilistic model of syntactic-semantic structure.

We compare the performance of the parser with the F-measures indicating the inter-annotator agreement in manual annotation ([22] and D6.2). If the output of the parser agrees with the manual data as much as the annotators agree among themselves, it suggests that the quality is as high as we can get from manual annotations.

The inter-annotator agreement on a random set of 100 sentences is given in the last row of Table 2. The parser's F-measure on predicate labelling of 51% when using filtered data is not far from the inter-annotator agreement of 59%. For argument labelling the parser trained on unfiltered data, which results in sub-optimal scores for predicate labelling, reaches an F-measure of 65% which is again not far from the inter-annotator agreement of 74%.

With respect to related work results are not strictly comparable, because the languages and annotations they adopt are not directly comparable to ours, and their methods have been evaluated on restricted test sets. But for completeness, note that our best result for predicate identification is an F-measure of 51% accompanied with an F-measure of 58% for argument labelling. Padó [18] reports a 56% F-measure

on transferring FrameNet roles, knowing the predicate, from an automatically parsed and semantically annotated English corpus. Padó and Pitel [23] report a best result of 57% F-measure for argument labelling given the predicate. Basili et al. [24] report 42% recall in identifying predicates and 73% global recall of identifying predicates and roles given these predicates. With respect to the task of training a semantic role labeller on projected annotations, Johansson and Nugues [25] have trained a FrameNet-based semantic role labeller for Swedish on annotations transferred cross-lingually from English parallel data. Evaluation is done on 150 translated example sentences. They report 55% F-measure for argument labelling given the frame.

4 Adaptation of an English parser to goal-directed spoken data

Within the framework of the CLASSiC project, we are working on goal-directed system-driven dialogues, where a system helps the user to fulfill a certain goal, e.g. booking a hotel room. Typically, users respond with short answers to questions posed by the system. For example *In the South* is an answer to the question *Where would you like the hotel to be?* Parsing helps identifying the components (*In the South* is a PP) and semantic roles identify the PP as a locative, yielding the following slot-value pair for the dialogue act: *area=South*. A PP such as *in time* is not identified as a locative, whereas keyword-spotting techniques as those currently used in dialogue systems may produce *area=South* and *area=time* indifferently.

Non-sentential utterances (NSUs) such as *In the South* are typical for goal-directed spoken dialogues. We would like to develop a syntactic-semantic parser that handles this type of data well. Statistical syntactic and semantic parsers need treebanks to train on. Current available data is lacking in one or more respects: Syntactic/semantic treebanks are developed on written text, while treebanks of speech corpora are not semantically annotated (e.g. Switchboard).

UNIGE adapted a semantic role parser to the domain of goal-directed speech by creating an artificial treebank from an existing text treebank [3]. We use a model that includes distributional models from both target (goal-directed spoken dialogues) and source domains (written text). We show that we improve the parser's performance on utterances collected from human-machine dialogues by training on the artificially created data without loss of performance on the text treebank.

4.1 Creating spoken data automatically

To construct a model of NSUs, we studied a subset of the data under consideration: TownInfo. This small corpus of transcribed spoken human-machine dialogues in the domain of hotel/restaurant/bar search is gathered using the TownInfo tourist information system [26] and contains 171 (approximately 5%) randomly selected utterances.

We define different types of NSUs based on the root label of the phrasal projection and define rules that allow us to extract NSUs (partial parse trees) from the source corpus.

By applying the extraction rules to the source corpus we build a large collection of both full sentences and NSUs. The distributions in this collection follow the distributions of trees in the source domain.

The small corpus of transcribed spoken human-machine dialogues is used to determine the target distribution of the different types of NSUs. We then sample from the extracted collection of both full sentences and NSUs following the distributions from the target domain and thus generate our artificial corpus. The

Training	Testing					
	TownInfo			PTB nonaug		
	Rec	Prec	F	Rec	Prec	F
PTB nonaug	69.4	76.7	72.9	81.4	82.1	81.7
PTB aug(+t)	81.4	77.8	79.5	81.3	82.0	81.7
PTB aug(-t)	62.6	64.3	63.4	81.2	81.9	81.6

Table 3: Recall, precision, and F-measure for the two test sets, trained on non-augmented data and data augmented with and without the target distribution component.

target distribution determines, for example, the proportion of NP NSUs that will be added to the artificial corpus.

We constructed our artificial corpus from sections 2 to 21 of the Wall Street Journal (WSJ) section of the Penn Treebank corpus [9] merged with PropBank labels [10]. We included all the sentences from this dataset in our artificial corpus, giving us 39,832 full sentences. In accordance with the target distribution we added 50,699 NSUs extracted from the same dataset. After the extraction we added a root FRAG node to the extracted NSUs¹ and we capitalised the first letter of each NSU to form an utterance.

4.2 Experiments

We trained three parsing models on both the original non-augmented merged Penn Treebank/Propbank corpus and the artificially generated augmented treebank including NSUs. We ran ontrastive experiment to examine the usefulness of the target distribution in the model by training two versions of the augmented model.²

The parsing model is the one proposed in [8], which extends the syntactic parser of [27] and [12] with annotations which identify semantic role labels, and has competitive performance.

In Table 3, we report labelled constituent recall, precision, and F-measure for the three trained parsers (rows) on the two test sets (columns). The results in the first two lines of the columns headed *TownInfo* indicate the performance on 150 transcribed utterances from TownInfo hand-annotated with syntactic and semantic annotation, that constitute our gold standard. The parser does much better when trained on the augmented data. The differences between training on newspaper text and newspaper texts augmented with artificially created data are statistically significant ($p < 0.001$) and particularly large for recall: almost 12%.

The columns headed *PTB nonaug* show that the performance on parsing WSJ texts is not hurt by training on data augmented with artificially created NSUs (first vs. second line). The difference in performance compared to training on the non-augmented data is not statistically significant.

The last two rows of the TownInfo data show the results of our contrastive experiment. It is clear that the three-component model and in particular our careful characterisation of the target distribution is indis-

¹The node FRAG exists in the Penn Treebank. Our annotation does not introduce new labels, but only changes their distribution.

²The model without the target distribution has a uniform distribution over full sentences and NSUs and within NSUs a uniform distribution over the 8 types.

pensable. The F-measure drops from 79.5% to 63.4% when we disregard the target distribution.

5 Conclusions

This task has achieved its main objectives. A prototype (D2.5) was completed on schedule. UNIGE has improved its domain-general syntactic-semantic parsing framework up to a point that it is one of the best parsing frameworks currently available. Results from the participation of UNIGE in the CoNLL-2008 shared task [4] indicate that synchronous parsing is an effective way of building joint models on separate structures. In [6], the model has been improved to better handle non-planarity in semantic dependency graphs. The flexibility of the probabilistic model, based on Incremental Sigmoid Belief Networks (IS-BNs), is demonstrated through the good performance on many languages, among which English, without the need for manual feature engineering from UNIGE's participation in the CoNLL-2009 shared task and its successful adaptation to the domain of goal-oriented spoken dialogue.

On the basis of the results discussed in Section 3 we can conclude that we have succeeded in successfully applying the parsing framework to French by using a knowledge-lean approach to semantic annotation transfer in combination with joint syntactic-semantic learning. The annotations provided by the joint syntactic-semantic parser for French, trained on a combination of automatically transferred semantic annotations and automatically parsed syntactic information, are of a higher quality than the automatically transferred semantic annotations on which it was trained. Moreover, the F-measure of the resulting annotations is only 8% points below the upper bound for predicate labelling and our best result for argument labelling is only 9% points below the upper bound.

6 Running the Prototype Syntactic-Semantic Parsers for English and French

The prototype consists of our synchronous model of syntactic-semantic parsing trained on 1) the English CoNLL data and 2) the automatically transferred data for French, described in section 3. It has been tested under Linux. First unpack `D2.5_prototype.tar.gz` by running

```
tar -zxf D2.5_prototype.tar.gz
```

Then either change to the directory `D2.5_prototype/English/`, for the English parser, or change to the directory `D2.5_prototype/French/`, for the French parser.

Regardless of which language's parser you choose, the procedure for running the parser is the same. First enter the directory `src/` and run

```
make
```

Then move back up to the previous directory (`D2.5_prototype/English/` or `D2.5_prototype/French/`), where there should now be an executable `synsem_parser`. From here you can run the prototype interactively by running the script

```
./interactive.sh
```

Simply type sentences, one per line, and the resulting parses will be displayed in the following format,

```
word-id word PoS head-id head-label predicate { argument-id argument-role }*
```

where “head” is the parent in the syntactic dependency graph, and the semantic role arguments for a given predicate are listed with the predicate. This script can also be run by piping sentences through them, as in

```
cat examples.snt | ./interactive.sh > examples.res
```

As illustrated inside the above scripts, the prototype can be invoked directly as

```
./synsem_parser -parsetext model/model1.par
```

In this mode, the output is produced in our “.ext” format, which can be converted to the official CoNLL 2009 format with

```
./scripts/ext2conll2009 model/deps.nonproj input_file output_file
```

The latter script also converts from projectivised syntactic dependencies back to the original non-projective syntactic dependencies used in the CoNLL 2009 data. Parameters, such as the beam width used by the syntactic-semantic parser for decoding, can be adjusted by changing the number which follows the parameter in `model/model1.par`, or by adding “PARAM=value” to the end of the above command, as in:

```
./synsem_parser -parsetext model/model1.par BEAM=10
```

Parameters can also be used to run the parser on files, with “TEST_FILE=*input_file*” and “OUT_FILE=*output_file*”.

Bibliography

- [1] J. Hajič, M. Ciaramita, R. Johansson, D. Kawahara, M. A. Martí, L. Màrquez, A. Meyers, J. Nivre, S. Padó, J. Štěpánek, P. Straňák, M. Surdeanu, N. Xue, and Y. Zhang. The CoNLL-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL-2009)*, June 4-5, Boulder, Colorado, USA, 2009.
- [2] P. Merlo and L. van der Plas. Abstraction and generalisation in semantic role labels: PropBank, VerbNet or both? In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 288–296, Suntec, Singapore, 2009.
- [3] L. van der Plas, J. Henderson, and P. Merlo. Domain adaptation with artificial data for semantic parsing of speech. In *Proceedings of NAACL*, 2009.
- [4] J. Henderson, P. Merlo, G. Musillo, and I. Titov. A latent variable model of synchronous parsing for syntactic and semantic dependencies. In *Proceedings of CONLL 2008*, pages 178–182, 2008.
- [5] A. Gesmundo, J. Henderson, P. Merlo, and I. Titov. A latent variable model of synchronous syntactic-semantic parsing for multiple languages. In *Proceedings of the thirteenth conference on computational natural language learning (CoNLL)*, 2009.
- [6] I. Titov, J. Henderson, P. Merlo, and G. Musillo. Online graph planarisation for synchronous parsing of semantic and syntactic dependencies. In *Proceedings of the twenty-first international joint conference on artificial intelligence (IJCAI-09)*, Pasadena, California, July 2009.
- [7] G. Musillo and P. Merlo. Accurate semantic parsing of the Proposition Bank. In *Proceedings of NAACL*, New York, NY, 2006.
- [8] P. Merlo and G. Musillo. Semantic parsing for high-precision semantic role labelling. In *Procs. CONLL*, 2008.
- [9] M. Marcus, B. Santorini, and M.A. Marcinkiewicz. Building a large annotated corpus of English: the Penn Treebank. *Comp. Ling.*, 19:313–330, 1993.
- [10] M. Palmer, D. Gildea, and P. Kingsbury. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31:71–105, 2005.
- [11] A. Meyers. Annotation guidelines for NomBank - noun argument structure for PropBank. Technical report, New York University, 2007.

- [12] I. Titov and J. Henderson. Constituent parsing with Incremental Sigmoid Belief Networks. In *Procs. ACL*, 2007.
- [13] I. Titov and J. Henderson. A latent variable model for generative dependency parsing. In *Proceedings of the International Conference on Parsing Technologies (IWPT-07)*, pages 144–155, Prague, Czech Republic, 2007.
- [14] M. Surdeanu, R. Johansson, A. Meyers, L. Màrquez, and J. Nivre. The CoNLL-2008 shared task on joint parsing of syntactic and semantic dependencies. In *Proceedings of CoNLL-2008*, Manchester, UK, 2008.
- [15] R. Johansson and P. Nuques. Dependency-based syntactic-semantic analysis with PropBank and NomBank. In *Proceedings of CoNLL-2008*, 2008.
- [16] D. Yarowsky, G. Ngai, and R. Wicentowski. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the International Conference on Human Language Technology (HLT)*, 2001.
- [17] P. Fung, Z. Wu, Y. Yang, and D. Wu. Learning bilingual semantic frames: Shallow semantic parsing vs. semantic role projection. In *11th Conference on Theoretical and Methodological Issues in Machine Translation (TMI 2007)*, 2007.
- [18] S. Padó. *Cross-lingual Annotation Projection Models for Role-Semantic Information*. PhD thesis, Saarland University, 2007.
- [19] R. Hwa, P. Resnik, A. Weinberg, C. Cabezas, and O. Kolak. Bootstrapping parsers via syntactic projection across parallel texts. *Natural language engineering*, 11:311–325, 2005.
- [20] F.J. Och. GIZA++: Training of statistical translation models. Available from <http://www.isi.edu/~och/GIZA++.html>, 2003.
- [21] P. Koehn. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of the MT Summit*, pages 79–86, Phuket, Thailand, 2005.
- [22] L. van der Plas, T. Samardžić, and P. Merlo. Cross-lingual validity of PropBank in the manual annotation of French. In *In Proceedings of the 4th Linguistic Annotation Workshop (The LAW IV)*, Uppsala, Sweden, 2010.
- [23] S. Padó and G. Pitel. Annotation précise du français en sémantique de rôles par projection cross-linguistique. In *Proceedings of TALN*, 2007.
- [24] R. Basili, D. De Cao, D. Croce, B. Coppola, and A. Moschitti. *Computational Linguistics and Intelligent Text Processing*, chapter Cross-Language Frame Semantics Transfer in Bilingual Corpora, pages 332–345. Springer Berlin / Heidelberg, 2009.
- [25] R. Johansson and P. Nugues. A FrameNet-based semantic role labeler for Swedish. In *Proceedings of the annual Meeting of the Association for Computational Linguistics (ACL)*, 2006.
- [26] O. Lemon, K. Georgila, J. Henderson, and M. Stuttle. An ISU dialogue system exhibiting reinforcement learning of dialogue policies: generic slot-filling in the TALK in-car system. In *Procs. EACL*, 2006.

- [27] J. Henderson. Inducing history representations for broad-coverage statistical parsing. In *Procs. NAACL-HLT*, 2003.