A hybrid approach to statistical machine translation between standard and dialectal varieties

Friedrich Neubarth [1], Barry Haddow [2], Adolfo Hernández Huerta [1], Harald Trost [3]

[1] Austrian Research Institute for Artificial Intelligence (OFAI)
[2] ILCC, School of Informatics, University of Edinburgh
[3] Inst. f. Artificial Intelligence, Medical Univ. Vienna, Austria
friedrich.neubarth@ofai.at, bhaddow@staffmail.ed.ac.uk, adolfo.hernandez@ofai.at, harald.trost@meduniwien.ac.at

Abstract

Using statistical machine translation (SMT) for dialectal varieties usually suffers from data sparsity, but combining word-level and character-level models can yield good results even with small training data by exploiting the relative proximity between the two varieties. In this paper, we describe a specific problem and its solution, arising with the translation between standard Austrian German and Viennese dialect. In a phrase-based approach of SMT, complex lexical transformations and syntactic reordering cannot be dealt with. These are typical cases where rule-based preprocessing of the source data is the preferable option, hence the hybrid character of the resulting system. One such case is the transformation between imperfect verb forms to perfect tense, which involves detection of clause boundaries and identification of clause type. We present an approach that utilizes a full parse of the source sentences and discuss the problems that arise with such an approach. Within the developed SMT system, the models trained on preprocessed data unsurprisingly fare better than those trained on the original data, but also unchanged sentences gain slightly better scores. This shows that including a rule-based layer dealing with systematic non-local transformations increases the overall performance of the system, most probably due to a higher accuracy in the alignment.

Keywords: statistical machine translation, hybrid approaches, preprocessing in MT, language varieties, dialects, parsing

1. Introduction

The standard paradigm of statistical machine translation (SMT) is tailored towards major languages with large bilingual and perhaps even larger monolingual text corpora at hand. Such mandatory prerequisites make it almost impossible to apply the same methods to less resourced (minor) languages, not to speak of dialectal varieties that most often completely lack written resources. On the other hand, less resourced languages and even more so dialects may be closely related enough to a resource-rich (major) language to exploit its resources in a sensible way. This offers new possibilities for the application of language technology methods to languages and varieties that were previously excluded from such a treatment. Instead of having huge corpora at hand for a particular dialectal language variety that offer themselves for machine learning techniques, methods can be developed to transform the input to SMT in such a way that it sufficiently resembles a resource-rich language. There are certain challenges with such an approach. For example, normally there is no authoritative orthography for dialects, which makes it necessary to develop a coherent standard for spelling and calls for methods to normalize the spelling of existing written texts. As for parallel (bilingual) resources, these may be even less common. However, the relative proximity between a standard language and its varieties makes it possible to gather parallel data and to establish SMT, despite data sparsity.

In addition, a resourced-rich language can be used as a ‘pivot language’ for translating a closely related less resourced language or variety into another major language. The SMT models of the pivot language are exploited by transforming the data of the less resourced language in such a way that it resembles the pivot language, a strategy that has been successfully applied to language pairs such as Macedonian and Bulgarian, or Bahasa Indonesia and Malayan (see section 2).

Another issue in SMT is that local dependencies can well be represented within phrase tables, but non-local ones usually cannot. Differences in syntactic structures that are reflected in different orderings on the level of terminal strings (words) pose specific problems. State-of-the-art MT attempts to resolve such reorderings by identifying the relevant sub-structures from tree-banks, and applying phrase-based SMT to the sub-trees while the tree structures are transformed according to the models of source and target language. Thus, syntactic reordering is captured over the tree structures. Generally, this can be neglected with closely related languages or varieties, since the syntax of the two languages is usually similar enough. However, there are still cases where syntactic reordering must be taken into account, even though the syntactic properties of the two varieties are almost identical. These arise when for a certain construction, there are two syntactic configurations available in the source language, but only one of them exists in the target.

In this paper, we will present such a case that appears in the context of translating Austrian German (AG), the standard variety, into a dialectal variety spoken in the capital, Viennese dialect (VD) (Schikola, 1954, Hornung,1998).[1] Most syntactic differences between these varieties can be attributed to morpho-syntactic properties and result in the different use of function words (e.g., relative clauses in VD often employ the indefinite pronoun *wås* ‘what’ in addition to the relative

[1] The work presented in this paper is based on the project ‘Machine Learning Techniques for Modeling of Language Varieties’ (MLT4MLV - ICT10-049) funded by the Vienna Science and Technology Fund (WWTF).

pronoun: *dea wås* ‘D-who what’). Nevertheless, these words generally appear in the same local context and can in principle be ‘learned’ by the phrase table. The phenomenon in focus is the lack of imperfect verb forms in VD (and most other Bavarian dialects).[2] Imperfect verb forms are synthetic, while the perfect used in VD is analytic, consisting of an auxiliary in the position of the finite verb and a past participle at the very end of the clause.[3]

The solution is apparently simple: one has to make the source look more similar to the target, i.e. imperfect forms are transformed into perfect (which is a legitimate option for AG). After this step of preprocessing, the phrase tables can be learned based on an input where the alignment is straightforward (auxiliary in finite position, participle at the end of the clause). Thus, the MT system has hybrid characteristics: data from the source language (AG) is preprocessed in a rule-based approach, incorporating linguistic knowledge about both varieties, and only after that, we can successfully apply statistical modeling. See Collins et al. (2005) for a similar approach to tackle the problem of non-local dependencies in German translated into English.

The most prominent reason for such a hybrid strategy lies in the fact that statistical alignment will not be able to identify the (remote) perfect participle as part of the verb group. Hence, only the imperfect verb form and the finite auxiliary end up in the phrase tables, and the lexical information of the verb, conveyed by the participle is lost in translation.

In the following section, we discuss some peculiarities of working with dialectal varieties and give a brief outline of similar methods that have been applied to other closely related language pairs. Section 3 provides some information about the bilingual corpus, and in section 4, we discuss in detail the rule-based component of our MT system. First results of combining this component with common SMT are presented and discussed in section 5.

2. Background

From a linguistic perspective, it has to be noted that dialects are never confined to a well-defined or delineated unique variety, also they are under constant influence of other varieties (most dominantly a standard variety), thus inherently subject to ample variation, synchronic as well as diachronic. Lacking (or resisting) standardization initiatives, reinforcement by education or public media and predominantly being confined to oral usage, dialects most often form a dynamic continuum between different varieties and speaker groups. Being defined by social group rather than geographical regions, the Viennese variety is a sociolect in the strict sense, where dialects in urban regions are generally associated with lower social classes (Labov, 2001). Moreover, speakers very often switch (or gradually shift) between their dialect and the standard variety, due to pragmatic reasons determined by the situation of the communication, the content, or to mark (pragmatic) emphasis. Therefore, the aim is to generate a consistent model of a dialectal variety that conforms to a stereotype of that dialect rather to incorporate all the variability.

Pairs of closely related languages (or language varieties) offer themselves to exploit the linguistic proximity in order to overcome the usual scarcity of parallel data. Nakov & Tiedemann (2012) take advantage of the great overlap in vocabulary and the strong syntactic and lexical similarity between Bulgarian and Macedonian. They develop an SMT system for this language pair by employing a combination of character and word level translation models, outperforming a phrase-based word-level baseline. The character-based SMT approach has been also used to process historical language, Scherrer & Erjavec (2013) follows this strategy to modernize historical Slovene words. Nakov & Ng (2012) propose a language-independent method to improve phrase-based SMT from a resource-poor language X1 into a language Y by exploiting the similarity of X1 to a related resource-rich language X2, by using bi-texts of the pair X2-Y. The proposed method is a hybrid approach of concatenation and combination of phrase-tables that are built by bi-texts X1-Y and X2-Y. Regarding MT of dialects, Zbib (2012) use crowd-sourcing to build Levantine-English and Egyptian-English parallel corpora; Sawaf (2010) normalizes non-standard, spontaneous and dialect Arabic into Modern Standard Arabic to achieve translations into English.

3. The corpus

Dialect speakers in Vienna often switch between the dialect and the standard variety, depending on the communicative situation, but also on the content that may invite to use a higher register. Text data with a bias towards the standard by virtue of standard orthography quite often also reflects such switching processes. In order to circumvent such biases, we carefully selected colloquial data of VD that are as authentic to the dialect as possible. The basic material consists of transcripts of TV documentaries and free interview recordings of dialect speakers. The transcripts (TR) were manually translated into both AG and VD in order to ensure that (rarely occurring) switchings to the standard variety do not end up in the target model, and to handle repetitions, truncations and some (uninformative) interjections.

(1)	TR:	kenn	ich,	jåjå	dån,	åiso	i	maan
	VD:	ken	i,	jå	dån,	åiso	i	maan
	AG:	kenne	ich,	ja	dann,	also	ich	meine
		know	I,	yes	then,	well	I	mean
		‘I know (it), yes then, well I mean’						

In an early stage, the task was to align these parallel sentences on a word-by-word basis, in order to simultaneously train a character-level translation model that would help to improve the alignment and generate lexical resources comprising morphology and morpho-syntactic features (PoS tags, grammatical features, such as number, person, tense, etc.). Usually the two translations (AG/VD) are syntactically very similar, with

[2] There are two exceptions: the auxiliary sein ‘to be’ and the two modals sollen ‘ought to’ and wollen ‘want’.

[3] A phenomenon with similar consequences for SMT is the lack of genitive case in VD. It is either replaced by dative, or – in possessive constructions – by a prepositional phrase (das Auto von der Schwester ‘the car of the sister’) or, with animate possessors, in a construction that does not exist in Standard German: the possessor in dative case, and a resumptive possessive pronoun (der Schwester ihr Auto ‘the sister-Dat her car’). These constructions will not be discussed in this paper.

little re-ordering and/or n-to-n correspondences. In addition, many corresponding words are 'cognates', meaning that they are lexically (and morphologically) the same in both varieties, with different phonology and spelling (e.g., AG '*also*' corresponds to VD '*åiso*' "thus/well"). The version of the corpus we used for the experiments described in this paper comprises 4909 sentence pairs with 39108 tokens for AG and 40031 tokens for VD. For further details regarding the development of a SMT system based on this corpus, cf. Haddow et al. (2013).

4. Rule Based Preprocessing

In VD, imperfect tense verb forms generally do not exist, but such forms quite often occur in AG. Sentences that express past tense with imperfect in AG have to use the analytic perfect tense in VD. Regarding AG, the choice between imperfect and perfect is more a matter of style than of meaning, perhaps due to the influence of Bavarian dialects spoken in Austria that always use perfect. Thus, transforming imperfect to perfect in the source language (AG) in order to match with the target (VD) still yields grammatical sentences.

The property that makes this task a real challenge is the verb second property of Germanic languages, which means that the finite verb in main clauses resides in a position next to an initial phrase, whereas in subordinate clauses it resides in its base position at the end of the clause (den Besten 1983). Crucially, the initial phrase in main clauses can be of any category that makes up a clausal constituent (noun/prepositional phrase, adverbial, but also a phonologically zero operator for yes/no questions, conditionals or discourse topics). In subordinate clause structures, the finite verb marks the end of the clause, phrases appearing to the right of it (and belonging to the same clause) have to be regarded as 'extraposed'. Consider the following (made-up, non-sense) example with a main matrix clause and the verb *sagen* 'to say/tell', and a subordinate clause headed by the conjunction *dass* 'that':

(2) Gestern sagte mir jeder,
Yesterday said-PAST to-me everybody
dass er seinen Esel schlug
whether he his donkey beat-PAST
'Yesterday, everbody said to me that he has beaten his donkey.'

The main verb (in its imperfect form) is in second position, the first being occupied by the adverbial *gestern*. The subordinate clause is extraposed, and the finite verb form *schlug* (again imperfect) appears at its end. Now, if we replace imperfect verb forms with perfect tense, the structures appear quite different:

(3) Gestern hat mir jeder gesagt,
Yesterday has-PRES to-me everybody said-PRT
dass er seinen Esel geschlagen hat.
that he his donkey beaten-PRT has-PRES
'Yesterday, everbody said to me that he has beaten his donkey.'

The finite auxiliary (*hat*) in the main clause is in 'second position', and the participle of the main verb appears at the end. In the subordinate clause the finite (modal) verb *soll* is at the end, the infinitive (*schlagen*) surfaces left adjacent to the finite verb. There are two further complications, the first concerning the auxiliary: not all verbs take the auxiliary of the base *haben* 'to have', some verbs such as *ankommen* 'to arrive' select an auxiliary form based on *sein* 'to be'. The second complication arises if the finite verb is a modal or the verb *lassen* ('let', 'have + V') with an infinitival complement, then an infinitive verb form occurs in the place of the participle main verb. This is called the IPP-effect (*infinitivus pro participio*). In addition, if there are more than two verbs in a subordinate clause, the order of them can be changed in certain ways, which are typical for specific dialects. In Bavarian dialects, the order main-verb > modal > finite-auxiliary (e.g., *lesen müssen habe* '*read had-to have*') turns out as main-verb > finite-auxiliary > modal. The rule-based transformation from imperfect to perfect tense is done in various steps; to illustrate how the procedure works, consider the list of individual steps

1. identify finite imperfect verb forms
2. identify the person and number features
3. generate the form of the appropriate auxiliary according to these features
4. generate the past participle of the main verb (or, if it is a modal, the infinitive form)
5. decide whether i) the clause is a main clause (with verb second) or ii) a subordinate clause
6. replace the finite verb with the auxiliary and
   - if i) find right boundary and place the participle (or the modal infinitive) there
   - if ii) place the participle before the auxiliary (or the modal infinitive after the auxiliary)

Provided extensive lexical resources for Standard German, some of these tasks are rather simple, in particular 1, 3, 4 and 6ii. Regarding task 2, person and number features are straightforward for all persons except 1P.Sg and 3P.Sg, since they use the same endings in imperfect verb forms (e.g., *sagte* '(I / s/he) said', *schrieb* '(I / s/he) wrote') but have different forms for the auxiliary (in present tense) used to form perfect tense (e.g., *habe* '(I) have', *hat* 's/he has'). This is done by checking the domain of the clause for a 1P.Sg.Nom personal pronoun (*ich* 'I'). If the features are identified, the generation of the appropriate forms (task 3) amounts to just looking them up in the lexicon (we use an FST compressed format for fast generation and lookup of full forms.) What is more difficult is to identify the appropriate base of the auxiliary. Transitive and many intransitive (unergative) verbs use the base *haben*, whereas a certain class of verbs (unaccusative) uses the base *sein.* A few verbs of this class are ambiguous. Some of these ambiguous verbs alternate between a causative (transitive) and an inchoative (intransitive) meaning, reflected also by the choice of auxiliary. Verbs of movement (especially if they have a goal argument, or a modification indicating directionality) generally belong to the class selecting *sein* (see Haider 1985), only in contexts where they convey a meaning expressing (physical) activity, they select *haben* in perfect tense. (E.g., *ich bin ins Zimmer getanzt* 'I danced into the room' vs. *ich habe die ganze Nacht getanzt* 'I danced the whole night long'). (See Diedrichsen 2002 providing a detailed analysis of the relevant (semantic) properties of these verbs.) For this 'proof-of-concept', we just collected a list of verbs that select *sein* or go with both auxiliaries depending on directionality/causativity and perform a lookup upon that list, but it would be preferable to retrieve this information from available lexical resources.

While task 4 is straightforward, task 5 and task 6i require sufficiently accurate information about the sentence structure, in order to determine the domain of the clause itself and to decide whether that clause has main clause or subordinate clause structure. Analyzing the string of tokens together with their PoS-tags does not provide sufficient information, so we decided to employ parsing in order to obtain structural information as well. For a first study, we used a standard parser for German, BitPar (Schmid 2004), which was trained on data from version 2 of the Tiger Treebank. The advantages of this parser – it employs the same set of PoS-tags (STTS) as we use in our lexicon and it is very efficient in terms of runtime as well as space requirements – are outweighed by the fact that the statistical model of the parser was trained on a news text corpus, whereas our data was collected from speech data. As a consequence, the parser did not deliver an output for all sentences (only for 584 out of 997, that is 58.5%), and where it did, the structures and labels were often incorrect in many ways. Re-training the data with colloquial data requires a large amount of syntactically annotated data – an unavailable option. Therefore the rule-based algorithm must yield greater robustness in order to determine the right clause boundary and to decide whether it is a main or subordinate clause structure.

The algorithm proceeds in the following steps: 1) perform a lookup on all terminal nodes, and if one is recognized as a finite imperfect verb form then assign these features regardless of the label coming from the parser. 2) for each finite verb, find the highest structural node that contains the verb but no other finite verb. This delimits the potential clause domain for a given finite verb. 3) determine the clause type using the following criteria: a clause is subordinated if i) it contains a subordinating complementizer (KOUS), ii) the functional label of the clause is RC (relative clause), iii) the verb is at the end of the clause domain, and the number of phrases preceding it is greater than one. Otherwise, the clause is considered a main clause. As one can imagine, the output of this algorithm is highly sensitive towards the parser output. Upon manual inspection, 129 out of 584 (22%) processed sentences were not grammatical and were therefore excluded from the training data.

5. Experiments with SMT

In this section, we report on some experiments using the data set described in section 3 and the set of preprocessed data as outlined in section 4 to build statistical machine translation systems, using Moses (Koehn et al., 2007).

The corpus is split into four sections, TRAIN, DEV, DEVB and TEST, where the first was used for estimation of phrase tables and language models, the second for tuning the MT system parameters and the third for testing during system development. The last was reserved for final testing. The three tuning and test sets contained 600 sentences each, while the rest (4909 sentences) were taken into the TRAIN set.

For SMT there are two options to build the phrase tables: on the word level or on the character level (using unigram or bigram character strings). While the word level models are useful to learn 'interesting' translations (i.e. different lexical items in source and target), it is highly affected by data sparsity, meaning that the number of out-of-vocabulary (OOV) words is very high. On the other hand, the character-level models are seriously affected by misalignments or by translations that involve different lexical items, but they can be very useful for the treatment of OOV words.

After observing the performance of word and character-level models in isolation, we decided to combine the two models into a *backoff* model. It uses the word-level translation wherever possible, but applies the character-level model for OOV words. A similar trait is presented in Nakov (2012), where the combination of a word and a character-model gave the best results when translating between closely related languages.

Earlier work on MT for closely-related languages (Vilar et al. 2007, Tiedemann 2009, Nakov and Tiedemann 2012), experiments with character-level translation models that are also built using phrase-based Moses, but allowing it to treat single characters or groups of characters as "tokens". In our backoff model, we use unigram character-level models that are trained on *cognates* from the training set to avoid training from "noisy" data, containing many German-Viennese word-pairs which either represent lexical differences, or are the result of bad alignments. To filter out the *cognates* from the statistically aligned data (using GIZA++, cf. Och and Ney, 2000), we used a function based on the Levenshtein distance between two candidates, log-normalized by length, where the two words are converted into a format similar to the output of the Kölner Phonetik algorithm (Postel 1969).

Using the word-level model as a baseline and the backoff model as the model relevant for testing, we can observe the following differences between models built on training data that has not been preprocessed and models built on preprocessed data (imperfect to perfect, reordering), tested on data from the DEVB set.

Model	original	preprocessed
baseline	63.28	64.01
backoff	68.30	69.10

Table 1: BLEU scores for DEVB sentences

For both, baseline and backoff model, there is a slight improvement on preprocessed training data. Note that the BLEU scores are relatively high compared to the typical values reported in the MT literature, reflecting the restricted vocabulary of the data set. Now, since the proportion between modified and unchanged sentences is rather unbalanced (only 39 of 600 sentences affected by preprocessing), it would be worthwhile to have a look on the results for the two different sets of sentences:

Model	modified		unchanged	
	orig.	preproc.	orig.	preproc.
baseline	49.67	56.71	64.26	64.68
backoff	55.75	61.02	69.13	69.84

Table 2: BLEU scores for modified and unchanged sentences of the DEVB set

Examining the performance on the modified (39) versus unchanged (561) sentences shows quite a big jump in BLEU on the modified sentences (as expected) but also a small improvement on the unchanged sentences. This shows that the performance of the SMT system gets better with preprocessed data due to an increased accuracy of the alignment, even though only a subset of sentences with imperfect (455 out of 997) could be successfully transformed into structures with perfect tense (and the verb forms in the right places).

6. Discussion and Outlook

It could be shown that preprocessing is a successful strategy for dealing with constructions that involve complex lexical transformations together with syntactic reordering. The reason why a phrase-based SMT cannot learn such a transformation lies in the non-local nature of the process. For MT between AG and VD, the necessary transformation from AG imperfect verb forms to analytic perfect (existing in both, AG and VD) is such a case. The improvement in performance not only affects sentences that display this construction, but also the overall performance yields better results due to an increased accuracy in the alignment. The bottleneck for preprocessing was the behavior of the parsing algorithm on our speech-based data, which provided either no output at all, or parses with distorted information about clause boundaries. Since what is needed is only the boundaries of major phrases (sentences and noun phrase level), we will attempt to replace the parsing component by methods that deliver just the needed output at the gain of higher robustness. Another issue is the inclusion of similar differences between AG and VD, in particular the lack of genitive noun phrases in VD (and other Bavarian dialects). Nevertheless, the presented proof-of-concept, showing that a hybrid architecture with a rule-based and a statistical MT component has advantages even under sub-optimal conditions makes us confident about the scalability this approach.

References

den Besten, Hans (1983). On the Interaction of Root Transformations and Lexical Deletive Rules. In: Abraham, W. (ed.) *On the Formal Syntax of the Westgermania*, Amsterdam: John Benjamins, pp. 47–131.

Collins, M., Koehn, P. and Kučerová, I. (2005). Clause Restructuring for Statistical Machine Translation. In: *Proc. of the 43rd Annual Meeting of the Association for Computational Linguistics* (ACL), Ann Arbor, June 2005, pp. 531–540.

Diedrichsen, E. (2002). Zu einer semantischen Klassifikation der intransitiven *haben*- und *sein*-Verben im Deutschen. In: Katz, G. et al. (eds.) Sinn & Bedeutung VI, Proc. of the 6th Annual Meeting of the Gesellschaft für Semantik, University of Osnabrück.

Haddow, B., Hernández Huerta, A., Neubarth, F. and Trost, H. (2013). Corpus Development for Machine Translation between Standard and Dialectal Varieties. In: *Proc. of the Workshop Adaptation of Language Resources and Tools for Closely Related Languages and Language Variants* of the 9th Int. Conf. on Recent Advances in Natural Language Processing (RANLP 2013), Sept. 13th 2013, Hissar, Bulgaria, pp. 7–14.

Haider, H. (1985). The case of German. In: J. Toman (ed.) *Studies in German Grammar*, Dordrecht: Foris, pp. 65–101.

Hornung, M. (1998). *Wörterbuch der Wiener Mundart*. ÖBV, Pädagogischer Verlag.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, Ch., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, Ch., Zens, R., Dyer, Ch., Bojar, O., Constantin, A. and Herbst, E. (2007). Moses: Open Source Toolkit for Statistical Machine Translation. In: *Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session*, Prague, Czech Republic, 2007.

Schikola, H. (1954). *Schriftdeutsch und Wienerisch.* Österreichischer Bundesverlag für Unterricht, Wissenschaft and Kunst.

Labov, W. (2001). Principles of Linguistic Change (II): social factors. Massachusetts: Blackwell.

Nakov, P. and Ng, H. T. (2012). Improving Statistical Machine Translation for a Resource-Poor Language Using Related Resource-Rich Languages. In: *Journal of Artificial Intelligence Research*, Vol. 44, pp. 179–222.

Nakov, P. and Tiedemann, J. (2012). Combining Word-Level and Character-Level Models for Machine Translation Between Closely-Related Languages. In: *Proc. of the 50th Annual Meeting of the Association for Computational Linguistics*, Jeju, Rep. of Korea, 8-14 July 2012, pp. 301–305.

Och, F. J. and Ney, H. (2000). Improved Statistical Alignment Models. In: *Proc. of the 38th Annual Meeting of the Association for Computational Linguistics* (ACL00), Hongkong, China, pp. 440–447.

Postel, H. J. (1969). Die Kölner Phonetik. Ein Verfahren zur Identifizierung von Personennamen auf der Grundlage der Gestaltanalyse. In: IBM Nachrichten, 19, pp. 925–931.

Sawaf, H. (2010). Arabic Dialect Handling in Hybrid Machine Translation. In: *Proc. of the 9th Conference of the Association for Machine Translation in the Americas* (AMTA), Denver, Colorado.

Scherrer, Y. and Erjavec, T. (2013). Modernizing historical Slovene words with character-based SMT. In: *Proc. of the 4th Biennal Workshop on Balto-Slavic Natural Language Processing of the 51th Annual Meeting of the Association for Computational Linguistics* (ACL 2013), Sofia, Bulgaria, pp. 58–62.

Schmid, H. (2004) Efficient Parsing of Highly Ambiguous Context-Free Grammars with Bit Vectors. In: *Proceedings of the 20th International Conference on Computational Linguistics (COLING 2004)*, Vol. 1, Geneva, Switzerland, pp. 162–168.

Tiedemann, J. (2009). Character-based {PSMT} for closely related languages. In: Marqués, L. and Somers, H. (eds.) *Proceedings of 13th Annual Conference of the European Association for Machine Translation* (EAMT'09), Barcelona, Spain, pp. 12–19.

Vilar, D., Peter, J.-T. and Ney, H. (2007). Can we translate letters? In: *Proc. of the 2nd Workshop on Statistical Machine Translation*, Prague, Czech Republic, ACL, pp. 33–39.

Zbib, R., Maldiochi, E., Devlin, J., Stallard, D., Matsoukas, S., Schwartz, R., Makhoul, J., Zaidan, O. and Callison-Burch, C. (2012). Machine Translation of Arabic Dialects. In: *Proceedings of NAACL: HLT 2012*, Montreal, Canada, pp. 49–59.