

# Enhancing Cross-border EU E-commerce through Machine Translation: Needed Language Resources, Challenges and Opportunities

Meritxell Fernández-Barrera<sup>1</sup>, Vladimir Popescu<sup>1</sup>,  
Antonio Toral<sup>2</sup>, Federico Gaspari<sup>2</sup>, Khalid Choukri<sup>1</sup>

<sup>1</sup>ELDA/ELRA, 9 Rue des Cordelières, 75013 Paris, France

<sup>2</sup>ADAPT Centre, School of Computing, Dublin City University, Dublin 9, Ireland

Email: {meritxell; vladimir; choukri}@elda.org, {atoral; fgaspari}@computing.dcu.ie

## Abstract

This paper discusses the role that statistical machine translation (SMT) can play in the development of cross-border EU e-commerce, by highlighting extant obstacles and identifying relevant technologies to overcome them. In this sense, it firstly proposes a typology of e-commerce static and dynamic textual genres and it identifies those that may be more successfully targeted by SMT. The specific challenges concerning the automatic translation of user-generated content are discussed in detail. Secondly, the paper highlights the risk of data sparsity inherent to e-commerce and it explores the state-of-the-art strategies to achieve domain adequacy via adaptation. Thirdly, it proposes a robust workflow for the development of SMT systems adapted to the e-commerce domain by relying on inexpensive methods. Given the scarcity of user-generated language corpora for most language pairs, the paper proposes to obtain monolingual target-language data to train language models and aligned parallel corpora to tune and evaluate MT systems by means of crowdsourcing.

**Keywords:** Statistical Machine Translation (SMT), e-commerce, user-generated content (UGC), domain adaptation, crowdsourcing

## 1. Lowering the Language Barrier to Strengthen Cross-border EU E-commerce

One of the challenges currently hindering cross-border EU e-commerce is the language barrier. Some studies have indeed shown that online consumers prefer home market suppliers and suppliers with whom they share the same culture and language (Cardona & Martens, 2014; Gómez-Herrera et al., 2014). The association eCommerce Europe (2015: 31) reports for instance the difficulties faced by merchants in culturally and linguistically diverse countries to reach the whole potential market. Ensuring the multilingualism of e-commerce sites thus appears to be key for the gradual development of cross-border EU e-commerce.

Nevertheless, given the costs associated with human translation and web localization services, entering the multilingual scene is not affordable for all types of enterprises. There is indeed still a percentage of enterprises for which setting up the infrastructure to undertake e-commerce is perceived as too costly compared to potential profits. In this scenario, a pan-European platform enhancing multilingual technologies for EU languages may reduce the costs of MT system development and thus become an economic asset for enterprises wanting to enter cross-border e-commerce. The benefits that could be reaped from this market enlargement may attract enterprises so far reluctant to such an investment. SMEs, which typically have limited resources available for language technology development, could be particularly impacted by the availability of such a platform. And given the weight of SMEs in EU's economy (28% of EU28 GDP, Muller et al., 2014: 14),

their increased participation in e-sales would have a non-negligible effect on EU e-commerce growth.

## 2. From Internationalization and Localization to MT: a Typology of E-commerce Textual Genres

E-commerce sites are made up of a variety of textual types that can be characterised along two main dimensions, as depicted in Figure 1: source (corporate-generated *vs.* user-generated content, UGC), and perishability (static *vs.* dynamic content). On the one hand, corporate-generated content encompasses text produced by the enterprise itself, describing, for instance, its domain of activity, its products and catalogues and selling procedures. This type of content is more stable, since it is meant to stay unmodified overtime. However, corporate-generated content also includes more dynamic textual types, such as press releases, news, event announcements, calendars and even blog entries. These textual productions are updated by new entries more often than static content. On the other hand, user-generated content is typically dynamic, featuring above all questions and opinions, but also more stable types of content such as blog entries or detailed product reviews. This indicates the existence of a certain heterogeneity within user-produced content, which includes carefully written (expert) reviews and also more fragmental and scattered opinions (Lu and Zhai, 2008), that can be found in discussion forums, microblogs (such as Twitter) and dialogue-based texts where users respond to each other's comments.

A third dimension, namely, conventionality, is highly correlated with the two previous dimensions: the prototypical corporate-generated static text will tend to

follow well-established conventions (e.g. in terms of orthography and style) and thus be more formal than a prototypical user-generated (UG) dynamic text, which will tend to use unconventional orthography and a freer writing style. The MT performance expectations, the challenges and the risks (both economic and legal) associated with the translation of each textual type are different and should thus be addressed specifically.

Static corporate content usually represents the first contact of potential customers with the enterprise, its quality and adequacy being therefore crucial. Inadequate product descriptions can bring about not only issues with client satisfaction, but also legal consequences if delivered products do not correspond to the descriptions on which the purchase decision was made. These legal hazards as well as the commercial risk of negative reputation place extremely high demands on the performance of translation systems, with localization services being the preferred option for translating this type of content. Tools which support localization abound. Many of them, such as the Poedit<sup>1</sup> desktop application, or the Pootle<sup>2</sup> and Weblate<sup>3</sup> web applications, support the Gettext Portable Object format, which is the *de facto* standard for multilingual websites.

Dynamic translations refer mostly to business or customer-generated content, which is produced regularly and in large quantities, so manual translation and even post-editing are far too costly options (Jiménez-Crespo, 2013: 89). This is why internationalization and localization of corporate sites tend to leave out dynamic content such as press releases or news items (Jiménez-Crespo, 2012: 154).

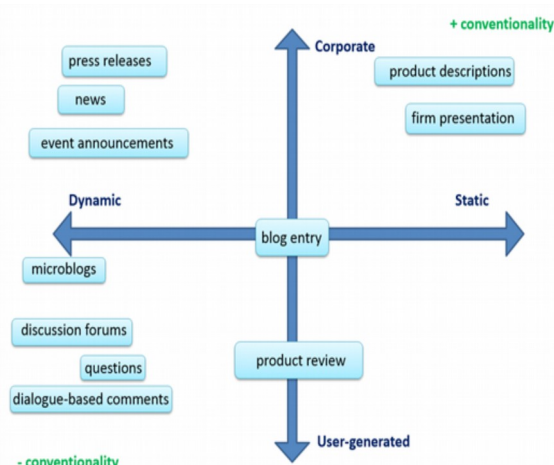


Figure 1: Typology of e-commerce textual genres.

Performance expectations are more relaxed with regard to dynamic content: corporate dynamic content (mostly press releases and news) does not usually touch upon core

product characteristics, and (translated) customer-generated content, if not endorsed by vendors, reduces the risk of litigation due to misleading product descriptions.

### 3. Challenges of UGC for MT: Unconventional Linguistic Features and Subtle Pragmatics

Specific challenges to be aware of have to do with the specific nature of dynamic content, especially the one produced by users. It has been shown that MT performance is improved if test data follow controlled language rules, because ambiguity and complexity are reduced (e.g. Aikawa et al., 2007). Since user-generated content (UGC) is at the opposite end of the spectrum with regard to controlled linguistic features, a poorer MT performance is to be expected. The non-controlled use of terms, the absence of (standard) orthographic conventions, loose grammatical constraints, and high dependence on the individuals' writing style, give rise to data sparsity issues affecting the performance of SMT systems, which present a poorer performance with informal genres (van der Wees et al., 2015a). This is why text pre-processing (i.e. text normalisation) is used as a method for improving SMT performance.

Different strategies have been deployed so far to enable automatic or semi-automatic text pre-processing for MT. The issue has been deeply studied in the framework of the ACCEPT project,<sup>4</sup> aimed at enhancing the translation of UGC in online communities. Seretan et al. (2014), for instance, describe a standalone pre-processing phase for defining correction rules targeting features of UGC negatively impacting SMT (e.g. use of slang, unconventional punctuation, ungrammaticality). Manually defined correction rules based on regular expressions can be either automatically applied (c.f. Jachmann et al., 2014) or they may require some form of intervention. Another strategy is to refine the SMT system by including a pre-processing step in which potential spelling errors are modelled (through a Confusion Network) and subsequently recovered by the decoder on the basis of a character *n*-gram language model (Bertoldi et al., 2010). However, both approaches have drawbacks. Whereas the first one requires manual intervention and is thus potentially slow and costly in terms of human resources, the second one entails high computational costs related to spelling-error modelling (Bertoldi et al., 2010: 418).

Beyond the structural particularities of UG text, further difficulties have to do with the translation of pragmatic nuances usually present in subjective and evaluative discourse. If those nuances are not accurately translated, the opinion contained in the original text may not be recoverable, or, worse, might yield, similar to lying, false implicatures, with significant consequences on the perlocutionary effects on the readers of the translated text (see e.g. Mejbauer, 2004, for an account on lying in relation with false implicatures, or Reiter, 1990, for a computational approach to avoiding conversational

1 <http://poedit.net/>

2 <http://pootle.translatehouse.org/>

3 <https://weblate.org/fr/>

4 <http://accept-project.eu/>

implicatures in computer-generated content). In this case, the translation would fail to meet users' expectations in an e-commerce context. The literature on social media analytics has proposed several techniques to leverage pragmatic content from online user interaction. The proposed solutions often rely on deep linguistic processing creating a semantic representation of text (e.g. Delmonte & Pallotta, 2011) and modelling argumentative structure (Pallotta et al., 2011) to capture meaning distributed over a conversation. Hybrid approaches to MT have in this sense been proposed to successfully handle the translation of semantic and pragmatic subtleties of UGC, e.g. the hybrid MT system PROMPT (Carrera et al., 2009).

#### **4. The Challenge of Domain Adequacy of Language Resources for E-commerce SMT**

One characteristic of e-commerce to be taken into account is the great variety of topics (as many as there are commercial sectors, such as travel, consumer electronics, clothing, etc.) and of textual genres (i.e. corporate dynamic text – news, events, customer evaluative text –). In terms of language processing this increases the risk of data sparsity. Mismatches between the training and the test data are likely to result in a deteriorated performance of the SMT system. This can apply both to lexical aspects (i.e. specific multiword expressions, collocations, out-of-vocabulary items, etc.) and to morphosyntactic particularities of the data, depending on the textual genre (e.g. ellipsis, inversion, lack of verbal agreement, lack of discourse markers, anaphoric or cataphoric constructions, etc.). An issue with regard to LRs needed for e-commerce MT is thus going to be how to ensure domain adequacy for a great variety of topics and genres or text types. Both topic (understood as the general subject matter of a document) and genre (understood as the non-topical text properties such as the style and register) are indeed considered crucial for domain adequacy (van der Wees et al., 2015b: 560-61).

The question to be answered is how SMT can deal with the heterogeneity existing among different types of UG text: for instance, whether an SMT system trained on blog entries (generally longer and to a certain extent following orthographic and grammatical conventions), is going to be equally suited to translating dialogue-like comment threads (containing much shorter and more informal texts). This question remains largely unexplored.

Only recently has the issue of genre been explored with regard to the SMT of UG content. In this sense, van der Wees et al. (2015b) have shown that topic differences do not completely explain translation performance differences across genres and that genre-specific translation errors are generally attributable to model coverage rather than to sub-optimal scoring of translation candidates.

### **5. Adaptation of an SMT System to the E-commerce Domain**

In this section we propose a methodology to translate e-commerce texts by means of SMT in a rapid and cost-effective manner. The aim of our approach is to enable e-commerce players, especially the ones that have so far been excluded from adopting MT, such as SMEs, to expand their online offer to additional languages in a cost-effective manner. We focus in particular on the case of building SMT systems to translate user reviews. As mentioned in Section 2, this poses considerably lower legal risks compared to the translation of enterprise-produced product information.

For example, if we consider the European e-commerce landscape, despite it consisting of a single market covering 28 states and over 500 million customers, according to the 2014 Eurobarometer,<sup>5</sup> the majority of e-commerce players (60%) do not sell in additional Member States beyond the one in which they operate. While there are of course other reasons for this issue that are beyond the scope of this paper (e.g. legal considerations, different VAT regimes, logistical complexity), one of the main factors has to do with the cost of tailoring products and/or services to new markets, and especially with the expenses associated with localization.

#### **5.1. Domain adaptation**

Our proposal falls in the area of domain adaptation, which consists of multiple techniques enabling easy porting of models trained on one domain to other domains (Banerjee, 2013). SMT systems require large amounts of training data (parallel and monolingual corpora, typically in the range of hundreds of thousand sentences), and smaller amounts of tuning data (normally in the range of hundreds to a few thousand sentences).

Our approach to domain adaptation focuses on acquiring and using domain-specific monolingual training data and parallel tuning data. We suggest leveraging these resources as follows:

- Training monolingual data. There are vast amounts of user review data available online that we can use to train the language model component of SMT systems.
- Tuning parallel data. Unlike other domains (e.g. parliamentary proceedings, user manuals), user reviews are usually not translated into other languages and therefore there is no availability of large amounts of parallel data for this data type. This does not allow us to adapt the translation model, because manually translating hundreds of thousands of sentences (the amount of data needed to train an SMT system) would clearly not be cost-effective, and would also slow down the entire MT system development process. However, we might strategically decide to manually translate a few hundred sentences in order to tune the SMT system on domain-specific

5 <http://ec.europa.eu/COMMFrontOffice/PublicOpinion/>

data. Clearly, it would be neither particularly expensive, nor time-consuming to translate a few hundred sentences and their impact on the SMT system's performance might pay off. To make this approach even more cost-effective we propose translating these sentences by means of crowdsourcing. Even though this crowd-based approach is relatively new, the effectiveness of crowdsourced tuning sets has already been demonstrated in some cases (Zbib et al., 2013), whereby crowdsourced translations have been shown to have the same value for SMT system development as professional translations, while considerably reducing costs. Crowdsourced data must be carefully monitored for integrity: preliminary tests carried out in connection with the research undertaken for this paper have shown that occasionally crowdworkers take shortcuts, e.g. using MT output to provide translations, which is undesirable for the final quality of the SMT systems. Our experience has shown that it is possible to implement effective mechanisms to discover cheaters and filter out their data, but unfortunately this quality control stage slows down the data generation process.

## 5.2. Experimental set-up

We now consider the specific case study of a company based in Spain that aims to expand its e-commerce operations internationally. The first additional language to consider in order to maximize the benefits of multilinguality across Europe and beyond would be English.

We first define the datasets used to build our SMT system, comprising data both from our domain of interest (user reviews) and from other domains, as follows:

- Parallel data for training. This is the data used to train the translation model. As mentioned in Section 5.1, most probably we would not find the large amounts of domain-specific (user reviews) parallel data ideally required for our language pair (English—Spanish), so we would use general-domain data, for which vast amounts are available in this language pair, e.g. from the WMT13 translation task.<sup>6</sup>
- Monolingual data for training. This is the dataset used to train the language model. We could crawl user reviews from many e-commerce websites, and fortunately there are effective resources already available for this purpose. We would use a collection of reviews from Amazon containing a staggering 82 million reviews (McAuley et al., 2015).
- Development data. This is a parallel (Spanish—English) dataset used to tune the weights of the different components of the SMT system on a

specific domain, with a view to optimising its performance.

- Test data. This again is a parallel (Spanish—English) dataset, in this case used to evaluate the quality of our SMT system. It is therefore important that this set is as similar as possible to the typical data on which the system will be deployed.

For development and testing we would use data from a corpus of user reviews in Spanish, the SFU corpus.<sup>7</sup> We would make sure that the reviews used for tuning and evaluation are disjoint. As for the size of the tuning and evaluation datasets, we would have e.g. 500 sentence pairs in each. While typically these sets use 1,000 sentence pairs or more, Pecina et al. (2012) showed that tuning sets of more than 400-600 sentence pairs do not improve translation quality (according to automatic evaluation metrics).

The motivation for using different datasets for training (Amazon) and tuning/testing (SFU) purposes is that, if we were to test on data from the same dataset as the training set (Amazon), the results could be criticised as being over optimistic and of limited application. As a matter of fact, for an e-commerce website to be able to replicate this approach in a realistic situation, it would need to have a huge amount of translated reviews already available to train on (which would be applicable only to a handful of very big e-commerce players). In contrast, by testing on a different dataset, our approach is more robust and its applicability open to any e-commerce player, e.g. SMEs with only very limited amounts of in-house reviews (to be used for tuning and evaluation), as they technically might be able to leverage review data from other e-commerce websites and use that to train an adapted MT system to translate their own reviews. This possibility notwithstanding, IPR issues will have to be cleared in advance, by negotiating the conditions for data reuse with the data holder.

While the use of Amazon reviews for training the language model is straightforward, we need to translate the SFU reviews, written in our source language (Spanish), into our target language (English) in order to use them for tuning and evaluation. To make this translation cost-effective, we suggest relying on crowdsourcing. We could use, for example, the CrowdFlower platform.<sup>8</sup> CrowdFlower allows users to configure jobs with a number of settings. We would use a subset of these settings with the aim of producing translations of as high quality as possible while the approach remains cost-effective. Details on the settings we would use follow, based on some preliminary tests that have yielded promising results:

- Performance level. CrowdFlower contributors are classified into three levels, according to their performance. Our jobs would be limited to level 3 (i.e. the highest-scoring) contributors, who at

<sup>6</sup> <http://www.statmt.org/wmt13/>

<sup>7</sup> [www.sfu.ca/~mtaboada/research/SFU\\_Review\\_Corpus.html](http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html)

<sup>8</sup> <http://www.crowdflower.com/>

the time of writing accounted for 7% of monthly judgements.

- Geography. One can select a set of countries from which workers are allowed to work on the job. We would limit the countries of workers to United States and Canada, to give priority to translations produced in American English (the variety of this language most frequently used in the Amazon data set employed for language modelling in our system).
- Language capability. One can restrict the contributors to work on the job by their language skills. One can select workers to be part of the so-called ‘editorial crowd’, defined as “highly competent in English spelling, syntax and grammar”. Our jobs would require workers from the editorial crowd, not to degrade translation quality.
- Speed trap. If this feature is activated, contributors are automatically removed from the job if they take less than a specified amount of time to complete a task. Our jobs would contain tasks of 10 translations each and a realistic time trap could be set to 150 seconds. Hence, if a worker were to take less than 15 seconds, on average, to translate each sentence in the task, they would be automatically removed from the job.

Once the settings are in place, we would prepare the task detailing the instructions for contributors, as shown in Figure 2. Subsequently, we would launch the task.



Figure 2: Screenshot of the translation task in CrowdFlower

Once the translation task is completed, we would check a random subset (e.g. 100 sentences) to make sure the data offered by the crowd is of acceptable quality, e.g. no crowdworker appears to have cheated by providing raw MT output to complete the translation task. We would keep track of the most frequent errors and assess how critical they are for the tasks of MT tuning and evaluation in an e-commerce scenario.

Next, we would train and tune the SMT system using the datasets described earlier on in this section. This step is straightforward using any state-of-the-art SMT toolkit,

such as Moses.<sup>9</sup> Finally, we would evaluate the performance of the system on the evaluation set, both with a representative set of automatic evaluation metrics (e.g. BLEU, TER and METEOR) and by means of a lightweight human evaluation. While MT evaluation with state-of-the-art automatic metrics is by definition cheap and fast, human evaluation should not be resource-intensive, to avoid wasting the benefits of the proposed approach to SMT system domain adaptation. Although we do not have conclusive findings on this yet, we would tentatively suggest that fluency evaluations of the raw MT output (performed by monolingual speakers of the target language), in addition perhaps to quick adequacy evaluations (e.g. based on coarse 5-point Likert scales) performed by bilinguals on a small subset of the MT output should be sufficient to check the quality of the SMT system performance. For a more targeted human evaluation, one could also focus on the accurate translation of high-frequency and particularly relevant key terminology and phrases that are salient in the specific domain of user reviews. Clearly, the need to limit the costs of human MT evaluation for budget-conscious SMEs exploring the option of cross-border EU e-commerce supported by MT would dictate the most suitable and realistic combination of evaluation methods, especially with regard to human approaches.

## 6. Conclusions and Further Work

In this paper we have suggested ways in which SMT can be successfully deployed to develop cross-border EU e-commerce, by describing current obstacles and identifying relevant technologies to overcome them. The work is particularly applicable to EU-based SMEs that wish to expand their operations by penetrating new multilingual markets and locales via e-commerce, not only within Europe, but also beyond. We have proposed a typology of e-commerce static and dynamic genres and discussed those for which the application of SMT appears more promising, pointing out the specific challenges involved in dealing with UGC.

In addition, the paper has specifically tackled the issue of data sparsity inherent to e-commerce settings and has reviewed state-of-the-art techniques to achieve the elusive goal of domain adequacy via adaptation. A robust workflow for the development of domain-adapted SMT systems based on inexpensive methods via crowdsourcing has been suggested that can be adopted by SMEs, given the very common problem of bilingual UG data scarcity in the e-commerce domain, e.g. with regard to text types such as user reviews. The proposed workflow has been presented through the use case of a Spanish SME wanting to use MT to disseminate user-created reviews of its products in English; this would minimize the legal challenges associated with the risks of using MT on sensitive commercial information. The proposed workflow has the advantage of relying on a careful use of crowdsourcing to generate the extra language data required for MT system training and development, leveraging existing data from neighbouring domains.

<sup>9</sup> <http://www.statmt.org/moses/>

As part of future work we intend to report in detail the results of implementing the workflow described in this paper, as tests are currently underway in a range of e-commerce sectors and involving multiple language pairs with significant business interest, both within Europe and worldwide. While initial results are encouraging, it appears that special care needs to be taken in deciding on the settings for the crowdsourcing jobs (e.g. screening procedures and speed trap for cheater identification and removal), in order to ensure the quality and reliability of crowd-generated data; some of these settings may have to be adapted depending on the countries where the crowdworkers are based and/or according to the language pairs involved in the tasks. Further tests will include MT quality evaluation components (with a mixture of automatic metrics and human approaches), to provide concrete indications in these areas that would be of great interest for the expansion of e-commerce. In connection to this, future work will test the performance of SMT systems developed with the workflow proposed in this paper on the translation of UGC, including, but not limited to, product reviews.

## 7. Acknowledgements

This research has been funded by FP7-ICT project MLI (Towards a Multilingual Data and Service Infrastructure, <http://mli-project.eu/>), grant agreement no.: 610951.

## 8. Bibliographical References

- Aikawa, T., Schwartz, L., King, R., Corston-Oliver, M. and Lozano, C. (2007). Impact of Controlled Language on Translation Quality and Post-Editing in a Statistical Machine Translation Environment. In *Proceedings of the MT Summit XI*, Copenhagen, Denmark, 10-14 September, pp. 1—7.
- Banerjee, Pratyush (2013) *Domain adaptation for statistical machine translation of corporate and user-generated content*. PhD thesis, Dublin City University.
- Bertoldi, N., Cettolo, M. and Federico, M. (2010) Statistical machine translation of texts with misspelled words. In *Proceedings HLT '10 Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 412—419.
- Cardona, M., Martens, B. (2014). Supply-side Barriers to Cross-border eCommerce in the EU Digital Single Market. JRC Technical Reports. Institute for Prospective Technological Studies. Digital Economy Working Paper 2014/3.
- Carrera, J., Beregovaya, O. and Yanishevsky, A. (2009). Machine translation for cross-language social media. [http://www.promt.com/company/technology/pdf/machine\\_translation\\_for\\_cross\\_language\\_social\\_media.pdf](http://www.promt.com/company/technology/pdf/machine_translation_for_cross_language_social_media.pdf)
- Delmonte, R., Pallotta, V. (2011). Opinion Mining and Sentiment Analysis Need Text Understanding. In V. Pallotta, A. Soro, E. Vargiu, *Advances in Distributed-based Retrieval Tools*. Volume 361 of the series Studies in Computational Intelligence. Berlin, Heidelberg: Springer-Verlag, pp. 81—95.
- eCommerce Europe (2015) European B2C E-commerce LIGHT Report 2015. Available at <https://www.ecommerce-europe.eu/facts-figures/free-light-reports>
- Gomez-Herrera, E., Martens, B. and Turlea, G. (2014). The Drivers and Impediments to Online Cross-Border Trade in the EU. *Information Economics and Policy*, 27.
- Jachmann, T., Grabowski, R. and Kudo, M. (2014). Machine-translating English forum posts to Japanese: On pre-editing rules as part of domain adaptation. In *Proceedings of the 20<sup>th</sup> Annual Meeting of the Association for Natural Language Processing (ANLP)*. Sapporo, Japan, January, pp. 808—811,
- Jiménez-Crespo, M. (2012). ‘Loss’ or ‘lost’ in translation: a contrastive genre study of original and localised non-profit US websites. *The Journal of Specialised Translation* 17, pp.136—163.
- Jiménez-Crespo, M. (2013). *Translation and Web Localization*. London, New York: Routledge.
- Lu, Y. and Zhai, C. (2008). Opinion integration through semi-supervised topic modeling. *WWW Conference*. Beijing, China.
- McAuley, J., Targett, C., Shi, J. and van den Hengel, A. (2015) Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 43—52.
- Mejbauer, J. (2004). Lying and Falsely Implicating. *Journal of Pragmatics* 37, pp.1373—1399.
- Muller, P., Gagliardi, D., Caliandro, C., Bohn, N. U. and Klitou, D. (2014). Annual Report on European SMEs 2013/2014 – A Partial and Fragile Recovery. European Commission Report, available at [http://ec.europa.eu/growth/smes/business-friendly-environment/performance-review/files/supporting-documents/2014/annual-report-smes-2014\\_en.pdf](http://ec.europa.eu/growth/smes/business-friendly-environment/performance-review/files/supporting-documents/2014/annual-report-smes-2014_en.pdf)
- Pallotta, V., Vrieling, L. and Delmonte, R. (2011). Interaction Mining: Making business sense of customers conversations through semantic and pragmatic analysis. In M. Zorrilla, J-N. Mazón, Ó. Fernández, I. Garrigós, F. Daniel, J. Trujillo (eds.) *Business Intelligence Applications and the Web: Models, Systems and Technologies*. IGI Global.
- Pecina, P., Toral, A. and van Genabith, J. (2012). Simple and effective parameter tuning for domain adaptation of statistical machine translation. In *Proceedings of the 24th International Conference on Computational Linguistics*, pp. 2209—2224.
- Reiter, E. (1990). The Computational Complexity of Avoiding Conversational Implicatures. In *Proceedings of Annual Meeting of the Association of Computational Linguistics (ACL)*, pp. 97—104.
- Seretan, V., Bouillon, P. and Gerlach, J. (2014). A Large-Scale Evaluation of Pre-editing Strategies for Improving User-Generated Content Translation. In

- Proceedings of the 9<sup>th</sup> International Conference on Language Resources and Evaluation (LREC)*. Reykjavik, Iceland, May, pp. 1793—1799.
- van der Wees, M., Bisazza, A. and Monz, C. (2015a). Five shades of noise: analyzing Machine Translation errors in user-generated text. In *Proceedings of the ACL 2015 Workshop on Noisy User-Generated Text*. Beijing, China, July 31, pp. 28—37.
- van der Wees, M., Bisazza, A., Weerkamp, W. and Monz, C. (2015b). What's in a Domain? Analyzing Genre and Topic Differences in Statistical Machine Translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7<sup>th</sup> International Joint Conference on Natural Language Processing (Short Papers)*. Beijing, China, July 26-31, pp. 560—566.
- Zbib, R., Markiewicz, G., Matsoukas, S., Schwartz, R. and Makhoul, J. (2013). Systematic comparison of professional and crowdsourced reference translations for machine translation. In *Proceedings of the Conference of the North American Chapter of the Association of Computational Linguistics*, pp. 612—616.