

Norway seen from a Portuguese vantage point – there and here

Diana Santos & Ana Rita Ferreira

Abstract: In this paper we provide a first look at how Norway is seen from Portugal through corpus, discourse and visual analyses, based on blogs written by Portuguese immigrants and visitors in Norway, and on an overview of all Norwegian works translated and published in Portugal. After presenting briefly LINGuateca's infrastructure for the computational processing of the Portuguese language, we introduce a small corpus of texts written about Norway by Portuguese bloggers, Plonorte, and analyse it according to their most distinctive lexical items, the colours used, and the emotions expressed. We are able to identify different behaviour in visitors and immigrants in these three aspects. We then visually analyse book covers of Norwegian books translated in Portugal, according to motif, colour, and visual weight in different cover parts, concluding that the depictions by Portuguese publishing houses are frankly positive towards Norway. Although both studies are preliminary, they point to an utterly positive view of Norway by Portuguese.

Keywords: corpora, statistical analysis, visual content analysis, blogs, translation, Portuguese, Norway

1. Introduction

In this paper we provide a first look at how Norway is seen from Portugal through corpus, discourse and visual analyses, based on blogs written by Portuguese immigrants and visitors in Norway, and on the overview of Norwegian works translated and published in Portugal.

Our goal is to assess through (quite different) empirical means the attitudes towards Norway – and things Norwegian in general – held by Portuguese people.

We used two different means to assess this, namely text analysis and visual content analysis, on two different datasets that we created for this purpose, namely the PloNorte (Through the North) corpus of blogs; and a collection of visually-annotated book covers of all Norwegian texts translated and published in Portugal (until 2016).

We will describe each analysis in turn, and end by suggesting some conclusions. But first, we will provide a small introduction to the context of Portuguese studies in Norway, that allowed us to perform the studies described here.

In fact, for historical reasons a large infrastructure for the computational processing of the Portuguese language was launched in Norway, precisely 20 years ago at SINTEF Oslo. This infrastructure, named *Linguateca* since 2002, in the form of a network with several nodes, has developed – and continues to develop to this day – a large set of

annotated corpora for supporting R&D in the Lusophone countries (see Santos (2014) for the corpus effort, and Santos (2015) for an historical commentary)¹. After the first author's joining the University of Oslo (UiO) in 2011, some of the research in Linguateca came to be done at UiO (see Ebeling et al., 2014), leading to (among others) the creation of the PANTERA corpus (Santos, 2017a, in press) and the STIG system (Santos, 2017b, Santos et al. 2017).

Both these resources were aimed at allowing translation studies in the Norwegian-Lusophone culture constellation, and implied the inventory of every translated text between the two languages – something which is obviously an open task since translation activity continues in both directions. While PANTERA as a parallel corpus is aimed at linguistic and cultural differences and similarities highlighted by translation, and is actively used by students of Portuguese at the University of Oslo; STIG as an information system was deployed precisely to allow considerations related to literature, edition, and reader reception, that require other kinds of information to be gathered. It was under the scope of the development of STIG for PANTERA that we amassed the relevant book covers (of the particular editions used in PANTERA) that made possible our second empirical investigation (over a subset of them).

2. What does a corpus of blogs mentioning Norway tell us?

We created manually a set of blog posts that discussed or mentioned Norway, written in Portuguese by Portuguese bloggers, and added author, gender, date, and whether the blogger was a) a tourist or traveller, b) an immigrant, c) someone else discussing Norway from a distance.

Examples from the blogs are shown in Figures 1-3.

¹ The main project is called AC/DC, *Acesso a corpos/Disponibilização de corpos* (“Corpus access / Making corpora available”).



TRIPEIRO NA NORUEGA

História de um tripeiro que se aventurou em rumar à Escócia e acabou na Noruega

QUARTA-FEIRA, 14 DE JULHO DE 2010

Multa!!!!

Sim, sim fui multado! Há sempre uma primeira vez...

Não desta feita não foi como o meu amigo Eric (Sueco ajuizado e que não diz mal dele a ninguém) que por conduzir a 107km/h, num local que não se pode ultrapassar os 90, pagou 620EUR... não, não comigo foi diferente, muito diferente.

Não é que me multaram por não comparecer a uma vacinação!!! Sim, não foram os 620 EUR, apenas uns 22 EUR, mas xiça eu nem sequer tinha sido notificado!

É assim, por cá, quem falha marcação é multado. Eis uma boa para recuperar a economia Portuguesa, cada tipo que falhasse uma consulta...pimba multa no gajo.

Ah não sabias? Informasses-te!

PUBLICADA POR TUGA NA NORUEGA À(S) 05:23 

ETIQUETAS: [CURIOSIDADES DA NORUEGA](#)

EI MALTA. QUE DIA É HOJE????



ETIQUETAS

[1 semana de trabalho na Noruega](#)
[2 semana na Noruega](#)
[3 semana na Noruega](#)
[A descoberta da Noruega](#)
[Akkerhaugen](#)
[Animais](#)
[Bach](#)
[Blues Festival](#)
[Coragem...](#)

Figure 1: Example of a blog from a Portuguese immigrant

ALGURES NO MUNDO

[HOME](#) [ABOUT](#)



Foram-me ao bolso e eu gostei. Noruega.

August 16, 2016 by Maria | [Leave a comment](#)

E pronto, há aquele dia em que a pessoa vai a um país para ser assaltada. Isso fomos nós este fim-de-semana quando decidimos ir à Noruega. Caro? Cara sou eu. Aquilo é para lá de qualquer balúrdio possível na terra. Ir ao mac, para duas pessoas, sem sobremesa, custou a pequena quantia de 23 euros. VINTE E TRÊS EUROS. E não me digam que é como a Suíça, que eu já lá fui, e não é. "Querida, quer respirar? Pague portagem", foi assim que me senti. Mas adorei. E, por isso, valeu a pena.

Mas vamos lá começar pelo princípio. Aí há uns meses, ou semanas, nem sei, decidi ir a algum sítio nestes 3 dias. Vi voos, e este estava razoável e pimba, marquei e pensei: "Ah, boa, vamos aos Fjords". Ahahaha. Ora, isto é mais ou

SEARCH

RECENT POSTS

- [Métodos contraceptivos que funcionam.](#)
- [Viagens pela nossa terra.](#)
- [Maravilhas de ser pai. Ou mãe.](#)
- [Há dois anos.](#)
- [África. Ou Moçambique.](#)

Figure 2: Example of a blog from a Portuguese visitor

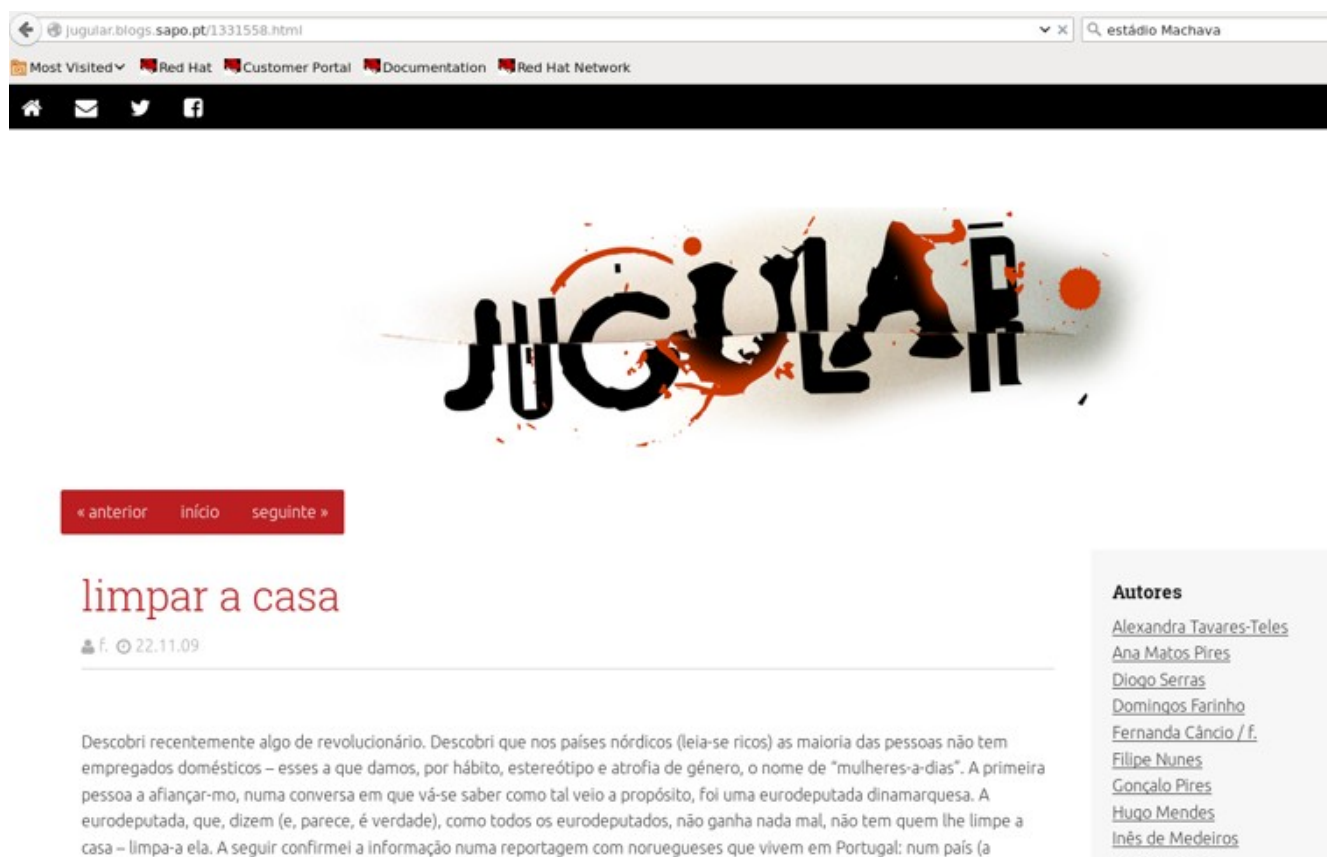
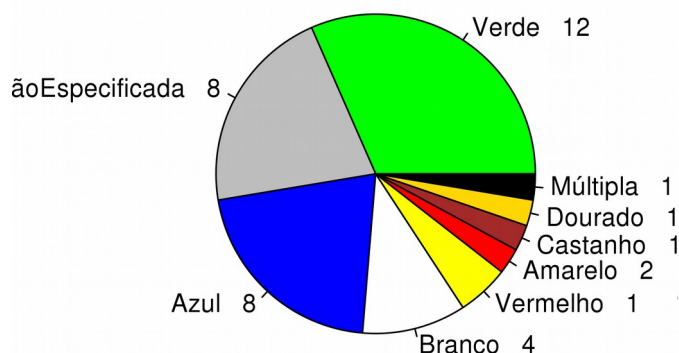


Figure 3: Example of a Norway-related blog post that neither belongs to an immigrant nor a tourist

The corpus, containing roughly 44,000 words from 38 blog posts, is available for inquiry from <https://www.linguateca.pt/acesso/corpus.php?corpus=PLONORTE>, together with the list of Internet sources. It contains 21,274 words written by tourists or travellers (here called “visitors”), 17,078 by people living in Norway (“residents”), and the remaining 5,363 by other people who wrote about Norway in their blogs (“others”). As all other corpora available through AC/DC, this corpus was syntactically annotated with PALAVRAS (Bick, 2000) and semantically annotated in several domains. In addition, it was subsequently humanly revised as far as colours and emotions are concerned. It is to these two domains we turn now:

Starting by the colours, our annotation distinguishes between real colours, and colours used in specific contexts like expressions, denoting race, human body, sports or political affiliation, or wine. In Figure 4 we considered only the real ones (i.e., we exclude metaphorical or conventional ones) and we show the colour groups used by the bloggers. The most striking conclusion is that visitors are much more visually impacted, and they describe Norway mainly in terms of green (both water and nature) and blue, but words like coloured or multicoloured (non specified colour) are very frequent as well.

Visitor, 38 in 21,274



Resident, 12 in 17,078

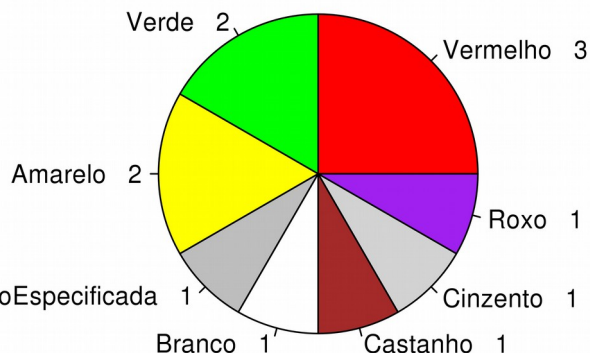


Figure 4: Colour use in the blogs of Portuguese visitors or immigrants

This contrasts with people who live in Norway, that use roughly three times less colours, and mainly for descriptions of concrete houses or products. (We should note that expressions translatable by “brown cheese” or “white spirits/wine” were considered product names – and marked as cor:original... otherwise brown, red, and white would have many more occurrences...)

Just for concreteness sake, let us share two occurrences of the lemma *branco* (white), the first in a tourist blog, the second in an immigrant blog:

par=46: A chegada a Bodo também se revelou muito bonita, com o barco a contornar dezenas de ilhéus e com as montanhas **brancas** a receber-nos. (Arrival at Bodo was also very beautiful, with the boat surrounding tens of small islands and with the **white** mountains receiving us)

par=97: Depois foi servida uma sopa horrível com umas bolas **brancas** que era peixe, batata e farinha prensado. (Then a terrible soup was served, with some **white** balls that consisted of mashed fish, potatoes and flour)

Doing the same to the emotion realm, and this time these categories are allegedly fuzzier, we were able to get the following “emotion clouds” for the three kinds of bloggers, in Figure 5, respectively visitors, immigrants and others.



Figure 5: Emotion clouds of the three different kinds of bloggers

What one can read from these data is that both visitors and residents employ most frequently the emotion *amor* (love, including friendship), in contradistinction to the “others”, who mostly employ *esperança* (hope) but discuss emotions like *fear* or *pride*. The two emotion domains that are more closely associated with being a visitor or an immigrant are, respectively, *surprise* and *saudade* (longing). Although this is not readable from the clouds, it is interesting that all bloggers (visitors, immigrants and others) have a high percentage of emotion words in their posts: 177, 154 and 51 (per thousand words: 8.3, 9.0 and 9.5). The fact that the highest emotion can be found in Portuguese writing about Norway from outside can be explained by saying that you have to love or hate a country to write about it from a distance.

We applied some statistical techniques to identify lexical preferences (which words were strongly associated with immigrants vs. visitors?), and used mutual information (for a critical view of a set of measures, see Kilgarriff (1996)). The words most associated with visitors came to be *mar* (sea), *comboio* (train), *cascata* (falls), *chuva* (rain), *ferry*, *km*, and *percorrer* (go through), while the words for immigrants are *comida* (food), *cerveja* (beer), *actividade* (activity), *quarto* (room), *beber* (to drink), *queijo* (cheese), *mala* (trunk) and *vinho* (wine). We believe that most of the words do not require comments, except for *mala* (trunk) which reminds one more of a traveller than a dweller. And this is in fact an interesting remark that should be made: Portuguese bloggers who live in Norway often blog about Norway because they are newcomers, and they are still either receiving friends and family or doing a lot of trips themselves.

In any case, one gets the idea that most Portuguese writers about Norway have a (sometimes, very) positive attitude about Norway, with maybe the exception of food and attitudes about alcohol. The most negative blogs in this (acknowledged) very small corpus are from a subset of the “other” writers.

Can this correlate, or have as one possible cause, the main view in Portugal through Norwegian literature? This is what we investigate in turn.

3. What do book covers tell us about attitudes towards Norway?

Book covers can tell us something about the book itself (and have often been used in reception studies in translation, see e.g. Bourdieu, Genette, cited in Brems & Pinto (2013) and Munday (2008)), but more than that, they are designed to appeal to their target groups. In that sense, they tell us something about the collective imagery of the potential readers and, at the same time, they help to create this imagery, since images can outlast what they represent.

Images are visualization of data and that was why we used visual content analysis. We discarded semiotic analysis, since we are not concerned with reading or interpreting each image by itself, but rather in relation to its peer-field and its origin-field. By peer-field we refer to the whole set of Portuguese book covers of the translations compiled in Pantera. By origin-field, we refer to the sample constituted by the respective Norwegian original book covers.

After creating a Web site (repository) to allow easy visualization of all covers², namely <http://noptbookcovers.weebly.com>, they were analysed according to the following variables: Colours; Motif and Highlights (or composition), each with their values

- Colour (values: blue, green, yellow, orange, red, pink, violet, pastels, black, white)
- Motif (values: typography, illustration, landscape, human figure (women, men, children))
- Highlights (values: text (title, author’s name), image)

Results for each variable are displayed in Figures 6-9.

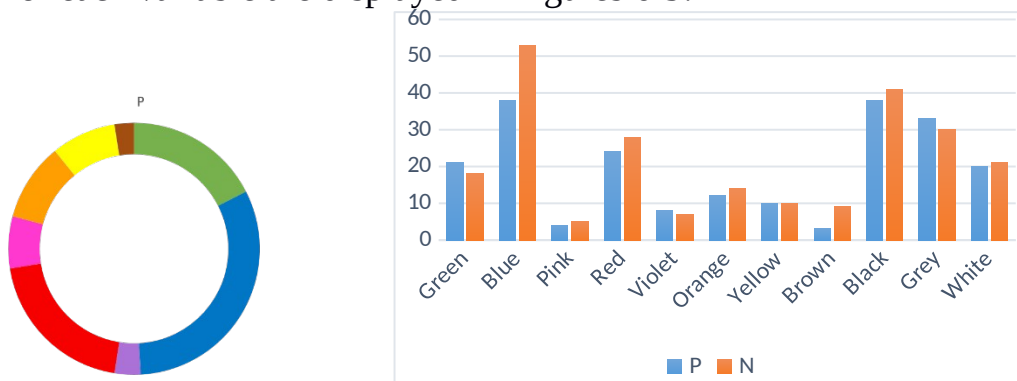


Figure 6: Colours used in the book covers of the Portuguese translations: palette (left) and comparison with the original covers (right)

In what concerns colour (see Figure 6), the palette is quite balanced in terms of warm colours and cold colours, despite the prevalence of blue tones. Interestingly, this palette is not so different from the one of the Portuguese visitors in Norway that we found in the PloNorte corpus. It had more green, but the other proportions were quite similar. Notice

² One should perhaps mention that the majority of the translated books are fiction, but not exclusively.

that there was not a significant difference between the peer-field and the origin-field in what concerns colour, but Norwegian covers do use more blue, brown and black.

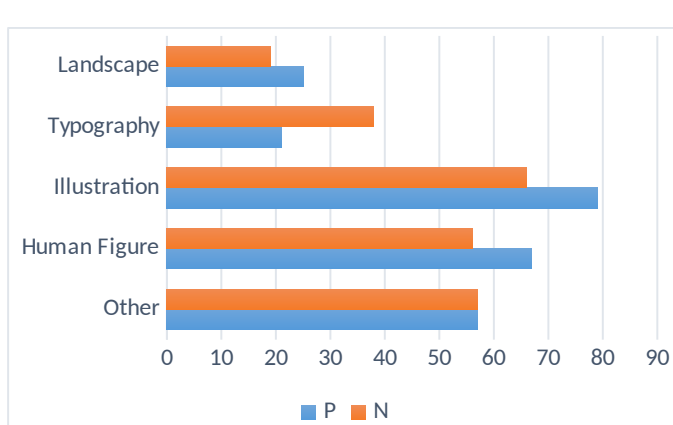


Figure 7: Motif in the book covers

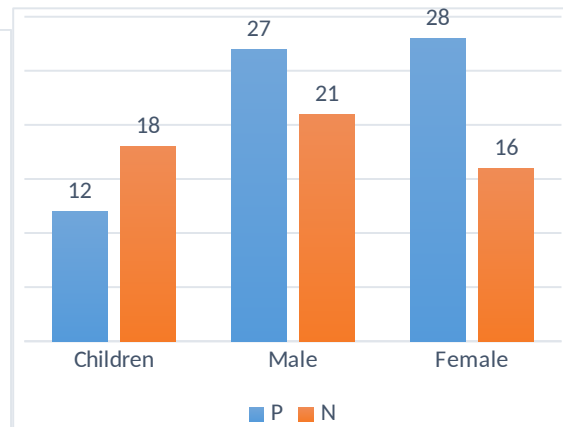


Figure 8: Gender presence in the book covers

As for Motif (Figure 7), Norwegian book covers use less illustrations. There are less landscapes and more human figures in the Portuguese translations. Boats were a common motif (they were grouped under the category “other”).

There were not as many children as in the origin-field group, but it was noticeable a more balanced gender representation in the Portuguese translations, with women being slightly more frequently depicted as Figure 8 shows.



Figure 9: Visual weight in the book covers. The two upper diagrams refer to origin field groups, while the lower pair of diagrams corresponds to translations.

Finally, Figure 9 compares composition in the two countries: there is an even distribution of visual weight between text and images in most of the book covers. But images do tend to be dominant in the cases where one of these elements is highlighted.

When the text prevails, the title of the book is more frequently spotlighted than the author's name.

But what can these data tell us about how Portuguese see Norway?

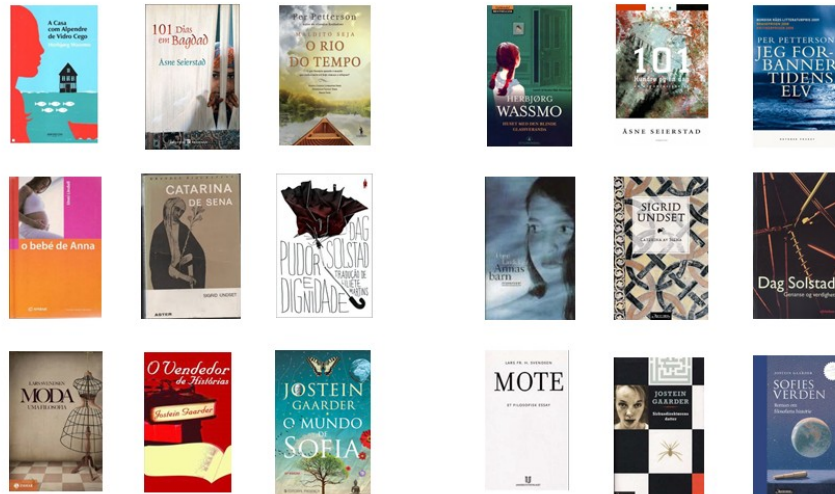


Figure 10: Examples of book covers in both countries: Portugal (left) and Norway (right)

Through the book covers, Lusophone readers do have a more positive image of Norway than the one that the Norwegian reader can find in her own country. Slightly warmer colour patterns, with less dark shades and less blue, characterize the covers of the Portuguese translations. Illustrations have here a prominent place and, in general, they provide a gentler and friendlier sensory experience than pictures, as Figure 10 attempts to show. Moreover, they are more effective in stimulating our imagination.

In addition, the number of book covers with typographic elements is much lower in the Portuguese translations, which means that these are visually less austere.

The fact that human figures are more likely to be found in Portuguese covers also helps to convey a sense of empathy, although the peer-field tops in what concerns images of children. And last but not least, and somewhat surprisingly, Norwegian literature in Portugal is visually more prone to gender equality than in its homeland.

The primary goal here was to obtain a quantified description of fields of representation. Inferences from visual content analyses are just a first step to further research. They can and should be supplemented by other methods that more adequately address layers of meaning. Visual content analysis, however, provides a solid starting point for such kind of projects.

This was just a preliminary endeavor, since there are two things that should follow up. First, the choice of the bookcovers themselves was relatively random, in that we did not (yet) have all the book covers for each source text and for all translations (we had only one per source text and one per translation). While this is not very critical for the peer-field, since there are not many re-editions of the translations, it is a clear issue for the origin-field, where especially for the classics, but not only, there is a multitude of editions of the same work. E.g. there are many editions of *Ett dukkehjem* in Norway (there were six before 1901³), and we were able to find eight for the detective novel *Elskede Poona* by Kari Fossum, with widely different covers (see Figure 11).



Figure 11: Three different covers of *Elskede Poona* by Kari Fossum

Given the well-known fact that choice of titles to translate is not a one-to-one affair (in this case, between Portugal and Norway) but is widely influenced by the choices in other (culturally surrounding) countries, and that translation has not always been direct, one has to take into account that other book covers may also have influenced the Portuguese choices. This is something that should and could be investigated for particular authors and/or publishers.

The second further analysis that could and should be performed is one before aggregation: namely, instead of comparing the distribution in the origin-field to the distribution in the peer-field as if they were independent, one might annotate for each work changes in colour (e.g. red to green), changes in motif (e.g. child to woman) and changes in highlights (e.g. author to title) and see whether other patterns emerged.

Finally, this preliminary analysis has been done without taking into consideration the publication date, which is an important factor, especially if we are comparing originals from the 1880s with translations from the 1970s (see Santos (in press) for a more detailed description of PANTERA's contents). To study the import of epochal fashions and printing technologies would be a relevant additional factor, something we hope to do in a later occasion.

3 According to http://ibsen.uio.no/VERK_Du.xhtml

4. Concluding remarks

In this text, one of our intentions was to show approaches of corpus linguistics and digital humanities that could be successfully employed in a study of attitudes between countries.

The two studies reported here were pilot studies to assess the relevance of possible follow-ups. Their presentation and the following discussion highlighted some issues that we would like to comment upon:

On the one hand, our search for blogs highlighted the almost complete disappearance of blogs, giving way to Facebook posts. This may imply that other kinds of materials should be amassed if one wants to do a larger study. On the other hand, the separation between the three kinds of bloggers can only be done by considering and reading every post, something which obviously does not scale up.

It was interesting to see more similarities from the two studies than would be expected, given that a priori there is not much in common between Portuguese publishers across more than one hundred years and bloggers about Norway in the 1990s. One could in fact also try to identify the view of Norway from newspapers in Portugal as a third possible indicator, in a similar vein as the study described in Ferreira (2015). One might thus check which adjectives and proper nouns co-occurred with unambiguous references to Norway, as Costa (2015) investigated about Japan.

Another, more focussed, empirical investigation would be to look at the translation of specific Norway-related cultural artifacts in the PANTERA corpus to see the implicit or explicit rendering of them by the translators into Portuguese, something we hope to do soon.

But so far we can conclude that the view of Norway from the South, at least from Portugal, is frankly positive.

Acknowledgements

We thank Linguateca's team for the infrastructure, FCCN for hosting the resources, the Department for Research Computing at USIT, the University of Oslo IT-department for computing facilities in abel, and ILOS for financial support in the development of both PANTERA and STIG. Some of the computations described here were performed on resources provided by UNINETT Sigma2, the National Infrastructure for High Performance Computing and Data Storage in Norway. We thank the organizers of the workshop for a very interesting venue.

Bibliography

Bell, Philip (2001), Content Analysis of Visual Images, in: Theo Van Leeuwen & Carey Lewitt (eds.), *The Handbook of Visual Analysis*. London: SAGE, 10–34.

- Bick, Eckhard (2000), *The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*, Aarhus: Aarhus University Press.
- Brems, Elke & Pinto, Sara Ramos (2013), Reception and translation, in: Yves Gambier & Luc van Doorslaer (eds.), *Handbook of Translation Studies*, Volume 4, 142–147.
- Costa, Luís Fernando (2015), O Japão visto de Portugal: Exploração usando um corpo de textos jornalísticos, in: Alberto Simões, Anabela Barreiro, Diana Santos, Rui Sousa-Silva & Stella E.O. Tagnin (eds.), *Linguística, Informática e Tradução: Mundos que se Cruzam. Homenagem a Belinda Maia*, Oslo: Oslo Studies in Language, vol. 7, n.1, 125–137.
- Ebeling, Signe & Grønn, Atle & Hauge, Kjetil Rå & Santos, Diana (2014), Corpus-based Studies in Contrastive Linguistics at the University of Oslo, *Corpus-based Studies in Contrastive Linguistics*, Oslo Studies in Language, Vol. 6, no. 1, 1–5.
- Ferreira, Ana Rita (2015), Lusophone discourses of the nation and the national, presentation at the seminar on *National identity and political symbolism*, UiO, 25th March 2015. <http://www.linguateca.pt/documentos/Ferreira2015.pdf>
- Kilgarriff, Adam (1996), Which words are particularly characteristic of a text? A survey of statistical approaches, in: *Proceedings of AISB Workshop on Language Engineering for Document Analysis and Recognition* (Sussex, April 1996), 33–40.
- Manghani, Sunil (2012), *Image Studies: Theory and Practice*, London: Routledge, 2012.
- Munday, Jeremy (2016), *Introducing translation studies: Theories and applications*. London: Routledge, 4th edition (1st edition, 2008).
- Santos, Diana (2014), Corpora at Linguateca: Vision and roads taken, in: Sardinha, Tony Berber & Ferreira, Telma de Lurdes São Bento (eds.), *Working with Portuguese Corpora*, Bloomsbury, 219–236.
- Santos, Diana (2015), Portuguese language identity in the world: adventures and misadventures of an international language, in: Elizaveta Khachatryan (ed.), *Language - Nation - Identity: The questione della lingua in an Italian and non-Italian context*, Cambridge Scholars Publishing, 31–54.
- Santos, Diana (2017), STIG, System for Translation Information in General: User manual and documentation". 2nd version, 30 October 2017, <https://www.linguateca.pt/Diana/download/STIGdocumentation.pdf>
- Santos, Diana (in press), PANTERA: a parallel corpus to study translation between Portuguese and Norwegian, to appear in *Proceedings of the XX Romanist Conference* (August 2017), BELLS.