Tracking Anglicisms in Domains by the Corpus-Linguistic Method—A Case Study of Financial Language in Stock Blogs and Stock Analyses

Anne Lise Laursen
Aarhus University
all@bcom.au.dk

Birthe Mousten
Aarhus University
bmo@dac.au.dk

Abstract - Lay investors and semi-professionals lean on professional stock bloggers and stock analysts for advice on stock investments; semi-professionals and professionals write about investments globally, and stock information has to be available in many local markets. Using the correct terminology for professionals, semi-professionals and to some extent lay investors is a challenge. Articles, blogs or translations involve a certain amount of specialized terminology. Against this backdrop, rub-off from leading English-language financial markets must be expected in local financial markets. In our case study, we compare the use of specialized Anglicisms in the Spanish and Danish markets, respectively. We discuss critical sense versus randomness in the adoption of Anglicisms in specialized financial contexts in the Danish and Spanish languages, thus arguing that the much coveted translator’s or a writer’s critical sense is not enough to make the right choices. Our corpus-linguistic tool can be a help in this specialized field.

Index Terms – Anglicisms, case study, corpus linguistics, critical sense, domains, financial terminology, LSP, professional blogs, randomness, specialized language, stock analysis, stock blog, stock exchange, technical writer, translation.

BACKGROUND AND RESEARCH AIM

Many lay investors and semi-professionals lean on professional stock bloggers and professional stock analysts for advice on financial matters. Writing professionally about the prospects of companies and markets in financial terms can be a challenge in relation to non-English-speaking lay people and semi-professionals. As experienced lecturers in specialized languages between English, Danish and Spanish, we have ascertained uncertainty in the assessment of acceptable choices between Anglicisms and local words in Danish and Spanish, respectively. This paper explores a way to use corpus linguistics to datamine through corpora to ascertain the extent and use of Anglicisms in Danish and Spanish stock blogs and stock analyses.

KEY CONCEPTS AND DEFINITIONS

Before venturing into our analysis of stock blogs and stock analyses, we shall address three concepts: Anglicisms, integration and corpus linguistics. Anglicisms: A general definition of Anglicisms is found in Oxford Advanced Learner’s Dictionary: “a word or phrase from the English language that is used in another language” [1]. However, for a word to be used in another language, according to Robinson, “it seems that borrowing is always involved, and if the word then remains in the second language it becomes a loan word.” [2]. Integration: Filipovic claims that receiving languages for Anglicisms “borrow English loans, adopt them as Anglicisms and subsequently integrate them into their vocabulary.” [3]. We agree with Filipovic in the borrowing–adoption–integration phases. Our perception of the integration phases is the following: in the borrowing phase the word is seen as a foreign element; in the adoption phase, the word is accepted in use, and in the integration phase—although still perceived as a loan—it is no longer seen as a foreign element, but as a logical choice of expression. We also find evidence of such thinking in Szubert’s work. He operates with “integrationsgrad” (extent of integration), defined as follows: “The transition from a foreign word to a loan word is not sharp and is only a metalinguistic differentiation. The delimitation between a foreign and a loan word is based on the extent of integration and recognition.” [4]. Gaudio also supports the language integration: “…through the quantitative and qualitative observation of how, and how frequently, an Anglicism is used, it is possible to determine the stage of incorporation of the etymon into the receiving language—the assumption being that hapax legomena, i.e. words occurring only once, are indicative of an earlier stage of incorporation, whereas recurring Anglicisms suggest later stages of incorporation.” [5]. In other words, not all
words go through the entire integration process; some loan words are rejected along the way and replaced by local words and phrases, whereas others may outcompete local words and become an accepted standard word.

The integration of Anglicisms comes in different forms. In our work with stock exchange corpora, we have evidenced incorporation through the use of direct loans, in which case words were imported and used directly, possibly with small grammatical inclinations, as hybrid loans, in which case loans were combined with a local word, and as false loans, in which case an English term was used locally and not in the English corpus in the same meaning. Corpus linguistics: Corpus linguistics can be defined as the study of language by means of machine-readable texts processed by an analytic software tool. Corpus linguistics is an online IT method which paves the way for a more systematic approach by including into a corpus linguistic program a number of texts from various sources within a domain, which are then analyzed on the basis of selected search criteria. To get to grips with randomness, the professional translator or communicator can use a corpus linguistic approach as a reliable tool for making an informed choice based on empirical studies of language use [6, p. 4].

LITERATURE REVIEW

Conventionally, studies of Anglicisms have been based on a word count in a fairly limited number of texts. For instance, Robinson has studied the use of English words in different sections of an Italian newspaper [2]. For the IT field, Rollason [7] refers to a number of articles on Anglicisms in French and Spanish as well as the book La Red by Cebrian. [8]. Rollason couples these with his own observations and thereby shows the randomness and a lack of logic in the choice of Anglicisms or local competitors in French and Spanish. One of many examples of such randomness and unpredictability is the French and Spanish use of the word ordinateur and ordenador, respectively, for computer, whereas the terms hardware and software are kept in Spanish but not in French. Conversely, Spanish has used Spanish words for server = servidor, browser = navegador and search engine = buscador. [7, p. 11] In 2015, Planchnon published results, which also showed the unpredictability of Anglicisms used in two newspapers online in Canadian media [9].

Anglicisms are used sporadically in specialized languages and are characterized by an extent of randomness, for instance in the sports domain of football (soccer), politics, etc. Our own observation from the football domain point in the same direction: The English term offside is adopted directly in Danish, but in Spanish the translation fuera de juego is used. Conversely, a penalty kick is translated directly into straffespark in Danish, but is adopted in Spanish as penalty. These examples are based on purely sporadic observations. The use of Anglicisms as a phenomenon, but not quantitatively evidenced, has been approached for different domains or eclectically for Danish by for instance Hye Christensen [10], Davidsen-Nielsen [11] and for Spanish by the abovementioned Rollason [7]—who confirms the unpredictability of Anglicisms in our research [12][13], where Spanish used more local equivalents than Danish when adopting Anglicisms within the SEO area.

OUR CLASSIFICATION OF ANGLICISMS

In our work, we operate with Anglicisms in the form of direct imports, such as cash-flow and shorting (in Danish), and as hybrid Anglicisms, which is defined as a combination of an English term and a local term, such as the Spanish term trading de intradia (=daytrading—in Danish this is a direct import in the form of daytrading.) Moreover, we include abbreviations and acronyms, such as the abbreviation EPS (Danish = EPS, Spanish = BPA), EBITDA (Spanish and Danish = EBITDA). These examples show the unpredictability of the occurrences of direct loans versus hybrid loans versus local competitors. Selected examples in the table below show the randomness or variation of the integration of Anglicisms in Danish and Spanish.

<table>
<thead>
<tr>
<th>English</th>
<th>Spanish direct loan</th>
<th>Spanish hybrid loan</th>
<th>Danish direct loan</th>
<th>Danish hybrid loan</th>
</tr>
</thead>
<tbody>
<tr>
<td>daytrading</td>
<td>trading de intradia</td>
<td>daytrading</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cash flow</td>
<td>cash flow (local competitor word: flujo de caja)</td>
<td>Cash flow (local competitor word: pengestrom)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bull market</td>
<td>N/A (local competitor: mercado alcista)</td>
<td>Bull-marked</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bear market</td>
<td>N/A (local competitor: Mercado bajista)</td>
<td>Bear-marked</td>
<td></td>
<td></td>
</tr>
<tr>
<td>leverage</td>
<td>Leverage (local competitor: apalancamiento).</td>
<td>gearing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If we take the example of leverage in English, it is translated into a false loan in Danish, gearing. In Spanish the local competitor apalancamiento, which by the way is a calque (direct translation of the concept of leverage), is preferred with the runner-up direct loan leverage. These...

1. SEO = search engine optimization
2. EPS = earnings per share; EBITDA = earnings before interest, tax, depreciation and amortization
3. beneficio por acción
examples again show that a translator’s or writer’s critical sense is of no help in an environment controlled by randomness and unpredictability.

THE CORPUS-LINGUISTIC METHOD

Koller has pointed out that corpus linguistics guarantees representative results, or at least more representative results than those generated by the analysis of isolated samples or by introspection alone [14]. Hultgren has called for corpus linguistics as a means to register Danish equivalents of English terms as well as the increasing use of English at Nordic Universities [15, p. 168]. Using corpus linguistics methodically, we have so far systematically researched the use of Anglicisms in the domain of search engine optimization [12] [13]. One conclusion was that Anglicisms are adopted in different ways in Danish and Spanish. Within the SEO domain, we found that Danish was more lenient in adopting Anglicisms immediately compared to Spanish, whereas the Spanish corpus was more stringent and creative in the search for local variants and alternatives. The Spanish corpus thus favored introducing Anglicisms by means of synonyms, code-glossing, code-mixing and exemplification to a greater extent than the Danish corpus. Such profound solutions and knowledge are often too time-consuming to find by manual datamining.

DATAMINING THE FINANCIAL DOMAIN

For the financial domain, especially two professional groups have to resolve the abovementioned language uncertainty. Firstly, communicators who write about financial issues, and secondly, translators into Danish and Spanish within the domain. One source of financial terminology is the auditors of a company and the auditing standards on corporate financial reporting that they lean on, for instance EBITDA or cash-flow. Another is the special language connected to the stock exchange jargon in particular, such as abbreviations like PE or specialized terms like shorting.

DATAMINING ANGLICISMS IN THE CORPORA

For tracking Anglicisms in the domain of financing, we have researched Danish and Spanish professional bloggers’ and analysts’ financial language use on banks’ and stock traders’ websites in the fall of 2014 and the spring of 2015. We have focused on the differences that emerge between the Danish and Spanish corpora. For such work, we show how small-sized corpora can help professional communicators solve their uncertainty questions in relation to their domain use of Anglicisms. For this purpose, we use a corpus-linguistic tool, AntConc.

Koller’s research described above is interesting for our method, because she is datamining for metaphors in her corpus. Because metaphors are not immediately found in a corpus unless you search for certain phrases or words directly, virtual corpus linguistics also has its limitations, although it is far better than manual datamining. Similar limitations and conditions apply when searching for Anglicisms in a corpus. A corpus-linguistic program cannot distinguish between common-core language, specialized language and Anglicisms in a corpus. In her search for metaphors, Koller has created “domains” for metaphors. A priori she knows that lots of metaphors group around words such as war, games, kill, struggle, etc., and she has used those words as her entrance to find metaphors in a virtual corpus.

Inspired by Koller’s a priori list—which method, by the way, was also used by Planchon [9]—we have dataminined our corpus with the purpose of tracking Anglicisms in stock blogs, stock analyses and financial texts by:

1) Searching via the AntConc program’s keyword list to find Anglicisms in the Danish corpus. This process brings forth specialized terminology and automatically filters away function words by comparing with words in a reference text. See appendix 1.

2) Adding familiar domain-specific Anglicisms from predetermined sub-domains to complete the a-priori list (financial ratios, accounting terms, stockbroker jargon and securities as well as market movements). These Anglicisms from the Danish corpus are checked in the Spanish corpus to see if they exist, to which extent, and how they behave and if they have a local equivalent. See appendix 2.

ANALYSIS APPROACHES

As shown just above, we have used two methodical approaches. For our corpus research, we have used the program AntConc, which is online freeware [16]. First we have applied the keyword facility to generate the keyword list from the Danish corpus and added our a-priori knowledge and observations to complete the a-priori list in the Danish corpus. The search for Anglicisms in the Spanish corpus has taken place by means of the AntConc Concordancer. Here we found evidence—or not—of Anglicisms in the Spanish corpus by comparing it to the Danish list. For local Spanish equivalents, we searched for either a calque of the English word (trend = tendencia) or by means of a word from the context (bulltrend = tendencia or tendencia alcista). Here, of course, it is important to have certain knowledge about

---

4 Price-earning

5 Tendencia is here the search word, which is used to find the collocation tendencia alcista.
the domain to be able to pin-point the appropriate equivalent (see appendix 3).

RESULTS FOR DOMAINS

In general, our research revealed a vacillation between Anglicisms, hybrids or local competitors in our two financial corpora in the following sub-domains:

Financial ratios

For the financial ratios, the Spanish corpus showed a mix of English and Spanish, which can be seen from the following example:

> Amadeus cuenta con un PER en el mercado de 24,5 veces para un BPA de 1,3 veces

where *PER* is an abbreviation for the English *price earnings ratio*, and *BPA* (*beneficio por acción*) is a local equivalent.

A parallel example from the Danish corpus was

> For 2015 ligger forventningen i øjeblikket på en EPS på 26,58, hvilket svarer til en price/earnings (P/E) på 11,8.

where both *EPS* and *price earnings (P/E)* are English terms, and whereby a certain stringency is demonstrated. Furthermore, some of the above abbreviations were backed up by full terms which appears from the Danish example above.

Accounting terms

For the accounting terms in particular, there seemed to be full agreement on adopting the English abbreviations, *EBIT, EBITDA, EV/EBITA* and the *pay-out ratio*. However, in Spanish the word order or hybrid constructions in the compounds were kept (for instance *margen EBITDA*).

Stockbroker jargon

When it comes to stockbroker jargon, the picture again turned to be varicolored. The actors in the market agreed on keeping *trading*, and *traders*, including *swingtrader* and *daytrader*—with Spanish, however, slightly vacillating between Anglicisms and hybrids:

* Es una plataforma muy adecuada para inversores de swing trading o traders de intradía.

Securities

For discussing securities, both languages agree on for instance *small-cap* and *mid-cap* for the corresponding indices for small and medium-sized companies.

Market movements

When discussing market movements, the Danish jargon kept the Anglicisms *bull market* and *bear market*, as well as *trend*, *uptrend* and *downtrend* and *upside* and *downside*. In these cases, Spanish rejected such Anglicisms in favor of the local hybrids with *alcista* and *bajista* (= rising and falling), combined with for instance *tendencia* (= trend) – see appendix 2.

The coexistence of Anglicisms and local competitors was only present to a small extent in the two corpora. This is quite interesting, because we saw a pronounced coexistence of Anglicisms and local competitors in our work on SEO corpora.

RESULTS FOR INTEGRATION

The number of occurrences gives us an indication of the use of Anglicisms in the two corpora, but it should be remembered that the compilation of our corpus has been on the fly which means that not necessarily all stock-market issues have been discussed in the blogs and market analyses compiled in this period. However, most of our findings are still salient on the basis of our small-sized corpus.

Firstly, there seems to be an unexpected divergence of the two key figures *EBIT* and *EBITDA* in the two languages. This is probably not due to the composition of the corpora or to the degree of integration, but to the accounting conventions of the market analysts in the two countries. In the Danish corpus, the term EBIT appeared 34 times and EBITDA 6 times, whereas the Spanish corpus showed results of 2 EBIT occurrences, but 33 EBITDA occurrences. We tentatively suggest that Danish accounting analyses tend to focus on EBIT as the central accounting term to compare annual accounts, whereas Spanish accounting analyses tend to focus on EBITDA as the central accounting term to compare annual accounts.

Secondly, for the actual integration, we found evidence of both zero occurrences in one language and an established use in the other language. An example of this is the English word *trend*, which had 12 occurrences in the Danish corpus and none in the Spanish corpus, where the local competitor *tendencia* was represented 241 times. Another interesting example is the term *blue chips*, which did not appear in the Danish corpus, but seemed to be common in the Spanish corpus with 30 examples without any local Spanish competitor. Terms like *emerging markets*, and composites with *shorting* are highly represented in the Danish corpus and totally absent in the Spanish corpus, where, in turn, the local competitors *mercados/países emergentes* and composites with *corto* seem well established as local coinages.

Thirdly, hapax legomena occurred several times. The terms *performance* and *bull* were found generally in the Danish corpus, but they had a borrowing status with only one occurrence in Spanish (hapax legomena), and were outnumbered by their local equivalents, cf. appendix 2. On the contrary, *RSI* (relative strength index), *ratio* and *yield* are equally represented and seem to be accepted Anglicisms in both languages.

CONCLUSION

In general, we found that our initial impression of randomness in connection with the choice between Anglicisms and local competitors has been documented by our corpus method approach. To a great extent, the Anglicisms from the Danish list were also present in the Spanish corpus. However, as Rollason points out, [7, p. 9] Spanish still proves to be rather successful in terms of finding local coinage of original English words. Still, the
amount of Anglicisms in the Spanish financial terminology indicates a certain willingness in this discourse community to lean on English terms when trying to understand the stock market.

The implication of our work here challenges the dogma that translators and communicators can solve language problems by using their critical sense. This is not the case for Anglicisms in specialized contexts. When freelance translators and professional writers have to embark on writing texts in local languages, their critical sense has to be combined with an analytical tool to establish the general use of Anglicisms in the domain in question. The reason for this is the level of integration, which may be found at different stages at any one time for any one Anglicism. This makes traditional mega-corpora vulnerable, because they do not have the time aspect, which can be documented by small, dynamic, on-the-fly, time-delimited corpora.

LIST OF REFERENCES:


Anne Lise Laursen is an emerica associate professor in Business Communication at Business and Social Sciences, Aarhus University, Aarhus, Denmark. Her research focuses on translation and corpus use in business contexts. She is a member of the Danish Association of Lexicographers.

Birthe Mousten lectures at Aarhus University, Aarhus, Aalborg University, Aalborg, and Copenhagen University, Copenhagen, Denmark. She also has taught courses for trade and industry. Her specialization is translation and text production. Her research areas are translation and localization as well as mediation.