Document généré le 2 mai 2024 17:03

Meta
Journal des traducteurs
Translators' Journal

META

Practice, Description and Theory Come Together – Normalization or Interference in Italian Technical Translation?

Silvia Bernardini et Adriano Ferraresi

Volume 56, numéro 2, juin 2011

Les corpus et la recherche en terminologie et en traductologie
Corpora and Research in Terminology and Translation Studies

URI : https://id.erudit.org/iderudit/1006174ar
DOI : https://doi.org/10.7202/1006174ar

Aller au sommaire du numéro

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)
1492-1421 (numérique)

Découvrir la revue

Citer cet article

Bernardini, S. & Ferraresi, A. (2011). Practice, Description and Theory Come Together – Normalization or Interference in Italian Technical Translation? *Meta*, *56*(2), 226–246. https://doi.org/10.7202/1006174ar

Résumé de l'article

Le présent article a pour objet la caractérisation de traits spécifiques de textes traduits : nous appuyant sur une expérience didactique, nous avons étudié l'emploi d'anglicismes dans des textes traduits ou non, dans le domaine de l'informatique. Le corpus utilisé à cette fin est composé de trois parties : des textes rédigés directement en italien, des textes sources rédigés en anglais, ainsi que les traductions de ces derniers. Les textes sources et cibles forment un corpus parallèle, tandis que les deux sous-corpus en italien forment un corpus comparable. Dans celui-ci, la fréquence de trois catégories de mots anglais a été comparée : emprunts directs, emprunts adaptés sur les plans morphologique et sémantique, et calques syntaxiques (pluriels terminant en –s). Le sous-corpus parallèle a ensuite été consulté pour réfuter l'hypothèse nulle selon laquelle les différences observées ne relèvent pas du processus de traduction. Les résultats de l'analyse quantitative, complétée par de scrupuleuses observations qualitatives, révèlent que les traducteurs se montrent plus conventionnels dans leurs choix lexicaux et normalisent davantage que les auteurs ; ceux-ci, au contraire, semblent plus enclins à accepter des interférences avec l'anglais, soit la langue véhiculaire dans le monde de l'informatique. L'article se termine par une discussion sur les implications de ces résultats au niveau méthodologique, descriptif/théorique et appliqué.

Tous droits réservés © Les Presses de l'Université de Montréal, 2011

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

https://apropos.erudit.org/fr/usagers/politique-dutilisation/

érudit

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

https://www.erudit.org/fr/

Practice, Description and Theory Come Together – Normalization or Interference in Italian Technical Translation?

SILVIA BERNARDINI
University of Bologna, Forlì, Italy
silvia.bernardini@unibo.it

ADRIANO FERRARESI
University of Naples "Federico II," Naples, Italy
University of Bologna, Forlì, Italy
adriano.ferraresi@unina.it

RÉSUMÉ

Le présent article a pour objet la caractérisation de traits spécifiques de textes traduits : nous appuyant sur une expérience didactique, nous avons étudié l'emploi d'anglicismes dans des textes traduits ou non, dans le domaine de l'informatique. Le corpus utilisé à cette fin est composé de trois parties : des textes rédigés directement en italien, des textes sources rédigés en anglais, ainsi que les traductions de ces derniers. Les textes sources et cibles forment un corpus parallèle, tandis que les deux sous-corpus en italien forment un corpus comparable. Dans celui-ci, la fréquence de trois catégories de mots anglais a été comparée : emprunts directs, emprunts adaptés sur les plans morphologique et sémantique, et calques syntaxiques (pluriels terminant en *–s*). Le sous-corpus parallèle a ensuite été consulté pour réfuter l'hypothèse nulle selon laquelle les différences observées ne relèvent pas du processus de traduction. Les résultats de l'analyse quantitative, complétée par de scrupuleuses observations qualitatives, révèlent que les traducteurs se montrent plus conventionnels dans leurs choix lexicaux et normalisent davantage que les auteurs ; ceux-ci, au contraire, semblent plus enclins à accepter des interférences avec l'anglais, soit la langue véhiculaire dans le monde de l'informatique. L'article se termine par une discussion sur les implications de ces résultats au niveau méthodologique, descriptif/théorique et appliqué.

ABSTRACT

This article aims at the characterization of specific features of translated texts. Taking a classroom experience as its starting point, the use of anglicisms in original and translated computing texts in Italian is examined. The corpus used for this purpose has three components: originals in Italian, comparable translations into Italian, and their English source texts. The frequency of three sets of English words – overt lexical borrowings, adapted borrowings and semantic loans, and morphosyntactic calques (plurals ending in *–s*) – is compared across the monolingual comparable subcorpus components. The parallel subcorpus is then checked to disprove the null hypothesis according to which observed differences are unrelated to the translation process. The results of the quantitative analysis, followed by careful qualitative observations, confirms that translators are more conservative in their choices and normalize more than writers, who seem to be more prone to interference from English as the lingua franca of the IT discourse community. Implications at the methodological, descriptive/theoretical and applied levels are discussed.

MOTS-CLÉS/KEYWORDS

universels de la traduction, normalisation, interférence, traduction technique, anglicismes
translation universals, normalization, interference, technical translation, anglicisms

1. Introduction

A widely quoted article by James S. Holmes (1972) described the multi-faceted nature of translation studies, a discipline in which descriptive, theoretical and applied concerns stand in a "dialectical [relation], with each of the three branches supplying materials for the other two, and making use of the findings which they in turn provide" (Holmes 1972: 78). This article seeks to provide, within a single piece of research, an illustration of this dialectical relation at work. The starting point is an applied one: how far is it appropriate to borrow and calque English words when translating documentation related to computer programming from English into Italian? This practical problem, which arose during a technical translation class, sparked a more in-depth investigation using corpus methodologies. The descriptive data gathered served to shed light on the current debate on norms and universals of translation in both theoretical and applied contexts. Finally, and again with reference to Holmes (1972), the article also contributes to a meta-reflection on the "methods and models" of translation studies (henceforth TS) in general, and corpus-based translation studies (CBTS) in particular.

In what follows we first briefly introduce our theoretical background: previous work is reviewed, in which *monolingual comparable corpora* were used in the search for *norms and universals* of translation, and the methodology's achievements as well as potential pitfalls are highlighted (2.1). The background section also describes the didactic setting in which the idea for this study originated (2.2). We then proceed to the study proper (3), first describing the underlying hypotheses and the corpus resources used, then illustrating the method and discussing the results obtained. In the last section (4) we consider implications at several interconnected levels: methodological, descriptive/theoretical, and applied.

2. Background

2.1. CBTS and the search for features of translated language

In a series of seminal articles dating back to the 1990s, and largely inspired by Toury's (1980; 1995) target-oriented approach to TS, Mona Baker (1993, 1995, 1996) advanced the idea that the role of corpora in TS was to elucidate "the nature of translated text as a mediated communicative event" (Baker 1993: 243). In pursuit of this objective, "universal features of translation" were to be identified, i.e., "features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems" (Baker 1993: 243). For this purpose, a new type of comparison was introduced, i.e., one in which translated texts were compared to original texts in the same language (rather than their source texts, as had hitherto been customary). A large body of empirical research has since employed *monolingual comparable corpora* (henceforth MCC) in the search for universal features of translation.[1] The new term *t-universals* has even been proposed to describe the hypothesized patterns emerging from the analysis of such corpora, distinguishing these *target*-oriented observations from *s-universals*, i.e., *source*-oriented observations such as those that could be derived from *parallel* corpus studies (Chesterman 2004).

Several hypotheses have been put forward about, on the one hand, potential translation universals (e.g., *explicitation*, *simplification* and *normalization*) and, on

the other hand, actual linguistic features associated with these universals that could be observed through corpus analysis. For instance, within a monolingual comparable corpus, a higher frequency of optional *that* with reporting verbs in translations has been suggested to be interpretable as evidence of explicitation (Olohan 2004), lower lexical density as evidence of simplification (Laviosa 1998), and lower frequency of words unattested in dictionaries ("coinages") as evidence of normalization (Williams 2005). Many other hypotheses about universals and associated features have been made and an exhaustive account would exceed the scope of this article. (The interested reader is referred to the excellent reviews by Laviosa 2002, Olohan 2004 and Mauranen 2008.) Suffice to say that the corpus-based approach to translation studies, and the MCC paradigm in particular, have produced a body of working hypotheses and empirical data that have enriched the discipline enormously – not least because of the methodological and theoretical discussions they have fuelled.

Leaving theoretical concerns aside, it is important to point out one fundamental methodological limitation of studies adopting the MCC approach. Since, by design, source texts are not included in these corpora, one must feel sure that the translated and non-translated texts differ *only* in the translation dimension. In other words, near-perfect comparability must be postulated, otherwise any observed differences could be due to unrelated or marginally related variables. If, for instance, Italian fiction texts translated into English were more high-brow than the English texts used for comparison, a more formal language would be observed in the translations than in the originals. Yet it would be wrong to conclude that greater formality is a feature inherent to the translation process. Observed differences would be related to translation through the much more tenuous link of preliminary norms (Toury 1995) that regulate decisions as to what texts are selected for translation into a specific target language, as opposed to those that are not.

Experience with corpus construction suggests that limited comparability between originals and translations in the same language might be the rule rather than the exception (Bernardini and Zanettin 2004, Mauranen 2008; on problems with the notion of corpus comparability in general, see Kilgarriff 2001). To get round this obstacle, Teich (2003) advocates the use of register-controlled corpora, on the assumption that a shared discourse community ensures closer comparability. While this is certainly sound advice, we believe that the ultimate test for a linguistic feature found in translated texts to be recognized as a feature of translation is its relationship to the source text. This is not to say that t-universals and MCC do not belong in CBTS, quite the contrary: MCC are arguably more versatile resources than parallel corpora, since the corpus-analytical techniques that can be employed with them are currently more sophisticated.

The framework we are advocating consists of a tripartite corpus structure in which the monolingual comparable component is used to identify quantitative differences across translated and non-translated texts (signaling potential t-universals), and the parallel component is used for the qualitative analysis of shifts accounting for the previously observed differences, through the painstaking, low-level analysis of parallel concordance lines. In Section 3 we provide more details about the corpus and see how the method works in practice. The next section moves back a step to present the didactic setting in which the research originated.

2.2. Didactic setting

The initial stimulus for this study came from an MA-level course in technical translation from English into Italian, jointly taught by the two authors at the School for Interpreters and Translators of the University of Bologna at Forlì (Italy) in the 2009/2010 academic year. The broad topic of the course was computing, for two separate reasons. First, in line with recent suggestions on best-practices in translation pedagogy (e.g., Olohan 2007, Koby and Baer 2003), the choice was guided by a concern with current work market demands: according to a 2009 study promoted by the European Commission on the state of the language services industry in the EU,[2] software and website localization accounts for approximately 25% of translation-related activities, and Information Technology ranks 4th among the fields of expertise most in demand by language professionals' clients. The second reason was a methodological one. Following Tim Johns' (1991) data-driven learning approach, students were encouraged to be active participants in the learning process, and to "identify problem areas, suggest hypotheses, and then test them together with their tutor who has the role of facilitator" (Laviosa 2006: 268); this has been suggested to better equip would-be translators with the skills and analytical tools they need to cope with real-life translation problems in relatively unknown Languages for Special Purposes (Laviosa 2006). As non-experts in computing ourselves, we were essentially on the same level as the students in terms of previous knowledge of specialized discourse conventions. We see this as an asset rather than a shortcoming: limited knowledge of the domain allowed us to engage with students in the learning process, and avoid the temptation to provide them with what Jean-René Ladmiral (1977, in Koby and Baer 2003: 211) called "*performance magistrale*," i.e., a teacher-produced standard towards which students should strive and against which their translations are evaluated. Using corpora in the translation classroom seemed an appropriate way of supplementing our limited subject knowledge while providing learners with key methodological and technical skills (for a recent collection on corpus use for teaching/learning, see e.g., Beeby, Rodríguez *et al.* 2009 and references therein).

Turning to the specific text type offered for translation in the classroom, we chose software documentation, and in particular documents known as "Perl pods": these are instructional texts that come as part of the standard distribution of Perl, a very popular programming language developed and maintained by a community of volunteers and distributed as a free and open source software.[3] Perl pods (or "pods," for short) serve as quick reference materials for programmers, providing them with short tutorials and answers on how to handle specific programming problems, e.g., the treatment of regular expressions in Perl. Figure 1 shows an extract from a pod document; notice the presence of "code text," i.e., instances of programming language alongside the actual text.

Figure 1

The pod source format (from: *perlretut.pod*)[4]

```
=head1 Part 1: The basics

=head2 Simple word matching

The simplest regexp is simply a word, or more generally, a string of
characters.  A regexp consisting of a word matches any string that
contains that word:

    "Hello World" =~ /World/;  # matches

What is this perl statement all about? C<"Hello World"> is a simple
double quoted string.  C<World> is the regular expression and the
C<//> enclosing C</World/> tells perl to search a string for a match.
The operator C<=~> associates the string with the regexp match and
produces a true value if the regexp matched, or false if the regexp
did not match.  In our case, C<World> matches the second word in
C<"Hello World">, so the expression is true. Expressions like this
are useful in conditionals:

    if ("Hello World" =~ /World/) {
        print "It matches\n";
    }
    else {
        print "It doesn't match\n";
    }
```

Pods were chosen for two reasons: because the subject matter is highly specialized, and because they are written in a special format (also called "pod"), and thus present students with the "technical" challenges they may face when translating documents in other special formats (an obvious parallel can be drawn with website localizers working on HTML code), a skill that is becoming increasingly essential for language professionals (Koby and Baer 2003: 212).

It is often the case with (open) software documentation that texts are originally written in English, and then translated by volunteers into different languages. Pods are no exception, and we know of at least two projects aimed at localizing the whole of English Perl documentation, one into Italian (called *Pod2it*),[5] and one into French.[6] For our purposes, this meant that a) we could count on expert support on the part of the *pod2it* project volunteers during the translation process, and b) we could provide students with a parallel corpus containing English originals and their translations. We also encouraged them to look for (roughly) comparable texts written originally in Italian, which would serve as additional evidence against which the validity of the solutions found in the Italian translated texts could be tested. As we shall see in Section 3.2, the same corpus design was adopted for the present study.

During discussion in class, students signalled a particularly problematic area in the translation of the texts, i.e., the choice between an anglicism and an Italian word as a translation equivalent for a given English expression. This is a much debated issue relevant to virtually all specialized discourses where the influence of the English language is strong. Discussing translation teaching in the areas of commerce and finance, Laviosa (2006) found the translation of anglicisms to be a sensitive area; similarly, Piqué-Angordans, Posteguillo *et al.* argue that:

> [t]he problem of borrowings in computer science in languages other than English has become a crucial one. […] There is no end in sight to the trend of admitting neologisms in information technology; they increase as computing science evolves and develops (Piqué-Angordans, Posteguillo *et al.* 2006: 222).

Through corpus consultation, students noticed contrasting tendencies in terms of the use of anglicisms on the part of translators and authors originally writing in Italian. Classroom discussion thus sparked the initial interest leading to the more systematic study presented in what follows.

3. Investigating anglicisms in Italian technical translation: a corpus-based study

3.1. Aims

Apart from their relevance for didactic purposes, differences in the use of anglicisms in translated and non-translated computing texts were hypothesized to also shed light on translation as a communicative event. Indeed, if translators were found to use more English words than Italian writers (as our students expected) this would be seen as evidence of interference from the source text. If, on the other hand, translators used fewer English words than the writers, a normalizing tendency could be hypothesized, i.e., a preference for the more normal or typical options afforded by the target language system. Previous work on normalization[7] has not focused specifically on foreign words, insofar as these are likely to be few and far between in the genres analyzed, e.g., literary texts (Kenny 2001; Englund Dimitrova 2004; Epstein 2010) and general-purpose web texts (Williams 2005). This is not the case in technical texts from the computing domain (Piqué-Angordans, Posteguillo *et al.* 2006), hence our interest and the decision to investigate the issue further.

But first catch your anglicism. Before the anglicisms can be compared across translated and non-translated corpora, they must be retrieved from the corpora, and this is easier said than done (Furiassi and Hofland 2007).[8] The Perl corpus is tagged with parts-of-speech (POS) and lemmatized, but of course there is no explicit tag identifying foreign words, nor is there any obvious heuristic that can be used to match non-Italian words using regular expressions. The compromise solution we found was to peruse wordlists and keyword lists manually, following up on the observations made in class. Thus there is no claim that *all* types of anglicisms are identified, but only that the same, uncontroversial types of anglicisms are collected from the translated and the non-translated subcorpora using one and the same method. Specifically, the three types of anglicisms focused upon here are, adapting Gottlieb's (2004) typology:

- overt lexical borrowings (i.e., new words);
- adapted borrowings and semantic loans (i.e., new verbs based on English roots with naturalized morphology and new homonyms of existing Italian verbs);
- morphosyntactic calques (i.e., *s*-plurals).

The method employed will be described in more detail in 3.3 below. In general, the research procedure is largely manual, and consists of observing differences in terms of the number of different anglicisms (types) as well as in terms of their frequencies (tokens) in the two Italian subcorpora. Cross-checks are then carried out in the parallel corpus to make sure that these quantitative observations can be attributed, beyond a reasonable doubt, to the translation process.

3.2. Corpus resources

The corpus used in the present study is a *composite bilingual* one (Laviosa 2006: 268), made of a parallel subcorpus (English ST and Italian TT) and a comparable (Italian TT) component. The parallel section includes the original English Perl pods distributed with Perl and their Italian translations produced within the Pod2it project, plus two short guides to programming in Perl (also written in pod format) for which we were able to find an Italian translation. The comparable component was gathered

performing various searches on web search engines and manually skimming through and evaluating the degree of "comparability" of the retrieved texts to the text type under scrutiny. In order for texts to be included in the corpus, they had to a) have a clear "instructional" function (i.e., provide guidance to other programmers), b) carry no indication of having been translated from another language, and c) be authored by Italians (as indicated by e.g., authors' bio notes).

Depending on the level of technicality of the topics covered, the texts in the corpus range from instances of "expert to initiates" communication (Pearson 1998: 38), where experts "will use the same terminology as they would use when communicating with their peers," but are also likely to "explain some terms which they believe to be […] inadequately understood by their readers" to instances of "teacher-pupil" communication, where authors "will use the appropriate terminology but will assume a much lower level of expertise" and where "[e]xplanations and definitions will be provided more frequently."

The downloaded texts were POS-tagged and lemmatized using the TreeTagger,[9] and indexed for consultation with the Corpus WorkBench (CWB),[10] resulting in three distinct subcorpora, i.e., PERLOREN (ORiginal ENglish), PERLTRIT (TRanslations into ITalian) and PERLORIT (ORiginal ITalian); the PERLOREN and PERLTRIT corpora were also aligned at sentence level using the CWB built-in aligner. Table 1 provides basic corpus information data.

Table 1

PERL: basic corpus data

	PERLOREN	**PERLORIT**	**PERLTRIT**
Tokens	298,346	305,537	321,405
Types	18,639	22,495	22,768
Texts	43	89	43
Writers	19	33	—
Translators/revisers	—	—	16

For the reasons outlined in Section 2.1, a critical aspect for purposes of this study was to achieve an adequate degree of comparability between the translated and non-translated texts included in our corpus. In this respect, Perl documentation affords very favorable, near-experimental conditions.[11] Compared to other investigations carried out within TS, in this case we are dealing with a neatly delimited topic (the Perl programming language), and a very homogeneous discourse community: both reference (original) texts *and* translations are drafted by area experts, not linguists, which allows us to factor out educational and professional backgrounds as potential intervening variables.

Of course, as in all manual corpus construction projects, any judgment we made during the selection of comparable texts was subjective, and thus questionable. As mentioned in 2.1, comparability itself is a problematic notion: the very fact that certain text types are translated into a target language may follow from a situation where domestic counterparts of the same text types do not (yet) exist. We believe, as Mauranen (2008: 37-38), that socio-cultural factors of this kind "are possible sources of systematic bias in […] databases, and impose limitations on their comparability," but also that "[the] search for generality cannot assume perfect homogeneity of the

research object." Bearing in mind these limitations, we thus take comparability to be more of a "guiding principle" in text collection (and analysis) than a straightforward, stringent requirement.

3.3. Method

For the monolingual comparable comparison of overt lexical borrowings (e.g., *package*, *script*) across originals and translations, frequency lists are first obtained for all word forms in the two Italian subcorpora. The two lists are then compared using the Log-Likelihood statistic (Rayson and Garside 2000), so as to identify those words that are significantly more frequent in one subcorpus than in the other. This might seem a somewhat unusual procedure: it is common practice in corpus linguistics to use corpora of the general language as baselines for filtering out frequent words from keyword lists (McEnery, Xiao *et al.* 2006). If a word is similarly frequent in a general-purpose and in a special-purpose corpus, then its frequency in the latter, however high, is not significant: the word in question is not a keyword. This is often a sensible choice, but not necessarily the only one. Depending on the reference corpus used, different words will be filtered out and different keywords will be selected for the researcher's perusal. Since the aim here is to identify words that are more frequent in translated than in original texts, it would not make sense to compare originals and translations in turn to a third (general-purpose) corpus. This operation would provide us with a list of the typical words in the two subcorpora (technical terms, genre-specific words etc.), but then the lists would have to be pruned of shared words, which are irrelevant for our purposes: if a borrowing were used by translators and authors alike, it would not be of interest to us. The pruning is instead done in one fell swoop if each of the two subcorpora acts in turn as a reference corpus for the other. Table 2 shows the top 20 entries from the two keyword lists.

TABLE 2

Top 20 keywords most typical of PERLTRIT when compared to PERLORIT and vice versa (Log-Likelihood order)

Typical of PERLTRIT		Typical of PERLORIT	
Word	**fq**	**Word**	**fq**
Unicode	297	nostro	355
consultate	256	riga	459
espr	175	quindi	361
potete	458	Apache	134
potete	244	istruzioni	233
vostro	415	cartella	107
package	357	infatti	160
versione	337	script	472
come	536	expression	104
byte	226	linguaggio	271
match	174	istruzione	298
consultate	107	abbiamo	369
CPAN	199	regular	103
Posso	201	array	882
maggiori	137	programma	734
dettagli	147	palline	73
undef	195	terminale	71
veda	126	contenuto	208
riferimento	308	vediamo	88
chiamata	319	puntata	76

The top 100 words (in Log-Likelihood order) with frequency equal to or greater than 5 are examined and all English-looking words are selected. The cut-off point is arbitrary, and only meant to keep the extent of manual analysis manageable. Parallel concordances (TT – ST) are browsed to confirm that the words thus identified are indeed used within "normal" Italian text (and not e.g., as part of a quotation, of untranslatable code text, etc.), that the higher frequency of a given English word in one subcorpus is indeed likely to be due to a preference by writers/translators for an anglicism rather than some other reason, and that solutions came from at least two texts written by different authors. In the case of translations this check was not deemed necessary since all translated texts are the product of at least two individuals, a translator and a reviser, working independently of each other.

Adopting a similar method but relying also on the morphosyntactic annotation performed on the corpus, two sets of verbs are identified that might seem to belong to the Italian lexicon,[12] but that in fact are either a) *adapted borrowings*, i.e., verbs containing an English root and an Italian suffix (e.g., splittare), or b) *semantic loans*, i.e., homonyms of existing Italian words with calqued meanings (e.g., transitive ritornare, ungrammatical in Italian and homonym of an intransitive Italian verb meaning *go back*; derived from transitive *return*, e.g., *return a value* ~ ritornare un valore). These words might easily escape a general search for English borrowings because at first sight they are indistinguishable from Italian words, therefore they are specifically focused upon through searches for verbs, followed by manual filtering. The cut-off point in this case is lowered (fq ≥2) since key-verbs are only a small subset of keywords. As in the previous case study, parallel concordance lines are then browsed for confirmation. Table 3 shows the top 20 entries from the key-verb lists used for the identification of semantic loans.

Table 3

Top 20 key-verbs most typical of PERLTRIT when compared to PERLORIT and vice versa (Log-Likelihood order)

Typical of PERLTRIT		Typical of PERLORIT	
Verb [lemma]	**fq**	**Verb [lemma]**	**fq**
consultare	383	eseguire	518
potere	1908	visualizzare	98
restituire	607	proporre	51
supportare	149	digitare	96
utilizzare	1022	ripetere	93
includere	221	dire	361
chiamare	441	individuare	84
fornire	241	analizzare	78
assicurare	73	servire	157
usare	929	contenere	606
impostare	209	memorizzare	126
codificare	65	valere	124
dovere	817	capire	108
funzionare	242	stampare	180
significare	159	premere	35
documentare	51	richiamare	62
omettere	97	assegnare	179
appropriare	37	ritornare	90
fallire	50	identificare	44
compattare	30	mettere	249

For the identification of morphosyntactic calques, we simply relied on a search for all words ending in *-s* in the two subcorpora, regardless of their parts of speech, followed by manual pruning.[13] In this case, no significance measure is needed since the overall number of candidates (<150) is small enough to allow for exhaustive manual inspection. Once again, the results obtained are checked through a careful analysis of parallel concordance lines. Table 4 shows the top 20 entries from the two lists of *s*-ending words.

Table 4

Top 20 most frequent s-ending words from PERLTRIT and PERLORIT

s-ending words in PERLTRIT		s-ending words in PERLORIT	
Word	**fq**	**Word**	**fq**
unless	85	warnings	69
this	60	Windows	56
bless	46	unless	54
exists	38	Mongers	38
threads	37	associates	31
Windows	36	keys	20
warnings	34	someclass	19
class	24	files	16
vars	24	class	14
extutils	22	tests	12
perlguts	19	pages	11
bytes	17	This	11
perlxs	16	threads	10
values	15	subroutines	10
keys	14	values	10
pos	13	this	9
statfs	11	tegels	8
fields	11	swords	8
attributes	10	dominus	8
signals	10	articles	8

3.4. Results

3.4.1. Borrowings

Several English words were identified among the top 100 keywords in the two lists. As can be observed in Table 5, more candidates were found in the translated corpus list than in the original one.

Table 5

Candidate English key-borrowings in translations and originals

English borrowings in PERLTRIT			English borrowings in PERLORIT		
Word	**Fq**	**LL**	**Word**	**Fq**	**LL**
package	357	178.4	script	472	131.0
match	174	148.2	expression	104	130.7
char	70	94.6	regular	103	123.2
filehandle	234	87.7	array	882	118.7
locale	115	83.7	overloading	59	75.0
require	112	83.3	print	920	54.1
unpack	60	72.3	reference	88	50.7
socket	102	66.9	matching	69	37.5
shift	208	66.9	Hello	24	34.1

local	140	65.6
buffer	62	63.7
point	55	54.9
record	70	54.4
long	50	53.4
pack	59	51.7
thread	282	50.5
Encode	36	48.6
pipe	99	46.5

These lists had to be pruned of words occurring mainly within (untranslatable) examples of code text (*char, filehandle*) and words that are more frequent in one list for topic-related issues (e.g., *locale* or *Encode* from the translated list, which refer to specific topics not covered to the same extent in original texts, and for which no equivalent Italian word exists). For the remaining candidates, an Italian equivalent was searched for in the Italian Perl subcorpora and on the web, to confirm that writers and translators using a borrowing *did* have a choice to use a native equivalent. In most cases we found that the English word was in fact the only option to express the concept in question, even in texts addressing a lay audience: this was the case with *socket, buffer, record, thread* and *pipe*. All these words have entries in the Italian *Wikipedia*, where they are defined and used as if they were standard Italian words. In some instances a literal translation is provided, but this is not always the case, cf. the case of *pipe*:

> Nei sistemi operativi una **pipe** è uno degli strumenti disponibili per far comunicare tra loro dei processi. (from Wikipedia: Pipe)[14]
>
> *In operating systems, a pipe is one of the tools that makes processes communicate with each other. (Translated by the authors)*

If there is no obvious alternative to the use of the term *pipe*, its presence in the corpus is indicative of topic rather than translator/writer preferences, therefore these cases are disregarded. Following this detailed analysis, we are left with the following borrowings (see Table 6), for which an alternative Italian word exists and is attested in this specialized field.[15]

Table 6

Real borrowings in translations and originals

English borrowings in PERLTRIT		**English borrowings in PERLORIT**	
Word	**Attested alternative Italian rendering**	**Word**	**Attested alternative Italian rendering**
package	pacchetto	expression	Espressione
		regular	Regolare
		reference	Riferimento
		Hello	Ciao

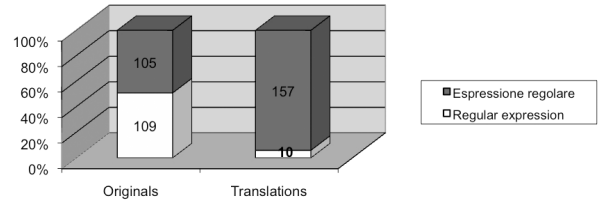
The situation is reversed with respect to the initial output (18 potential borrowings in translations vs. 9 in originals). The only borrowing seemingly favored by translators over the Italian alternative term is *package*, which occurs 357 times in translations and 81 times in originals. However, the Italian equivalent pacchetto is also much more frequent in translations than in originals (453 vs. 96 solutions), sug-

gesting once again that we are observing a topic-related difference (translated texts deal with *packages*/pacchetti more than native Italian texts).

Moving to the non-translated subcorpus, the first two key-borrowings are in fact part of the same phrase, namely *regular expression*/espressione regolare. If we add the frequencies of the English borrowing and of its Italian alternative, we get the number of times that the notion is referred to explicitly in the two subcorpora: 167 times in the translated subcorpus and 214 times in the non-translated subcorpus. In the overwhelming majority of cases (94%) translators opt for the Italian term, while writers equally prefer the English borrowing or the Italian equivalent (51% vs. 49%; see Figure 2).

FIGURE 2

***regular expression* vs. espressione regolare in originals and translations**



Figures 3 and 4 present data for *reference*/riferimento and *Hello*/Ciao that confirm the pattern observed in the case of *regular expression*.[16] Where an Italian alternative is available, translators show a very clear preference for it (over 90% of the total). Writers also use the Italian term in a majority of cases, but in over 30% of the total they opt for the English word, again showing a less clear-cut preference for the Italian word over the English borrowing.

FIGURE 3

***reference* vs. riferimento in originals and translations**

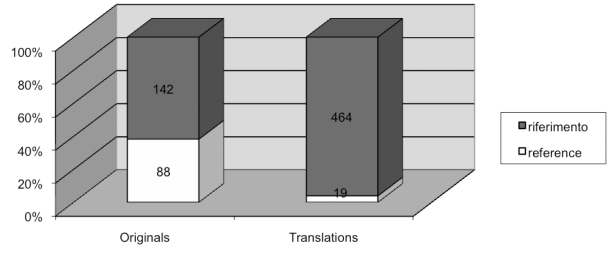
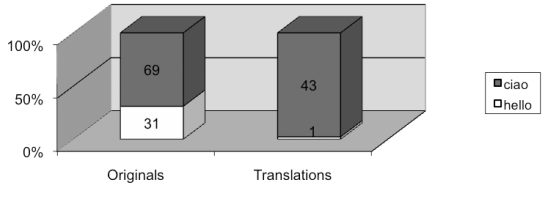


FIGURE 4

***Hello* vs. Ciao in originals and translations**



3.4.2. Adapted borrowings and semantic loans

Turning to the analysis of verbal borrowings and loans, manual inspection of the lists of the most "typical" verbs when comparing the two corpora (in the sense outlined in Section 3.3) allowed us to isolate, with the help of standard lexicographic resources for the Italian language, the verbs 1) whose root is derived from an English word, but is morphologically naturalized through the addition of an Italian verbal suffix (adapted borrowings); and 2) that exist in Italian but acquire a new sense and new collocational/colligational patterns derived from those of an English cognate. After this manual pruning, we were left with 9 verbs, shown in Table 7; notice that for adapted borrowings the counts refer to word forms,[17] while for semantic loans they refer to the frequency of lemmas.

Table 7

Verbal loans and borrowings in originals and translations

	Adapted borrowings			Semantic loans		
	Word [form]	**Fq**	**Fq / 1,000 words**	**Word [lemma]**	**Fq**	**Fq / 1,000 words**
PERLORIT	istanziare	4	0.013	ritornare	90	0.294
	istanziamo	2	0.006			
	cicliamo	2	0.006	processare	26	0.085
	cicla	2	0.006			
	splittare	3	0.009	valorizzare	8	0.026
PERLTRIT	–	–	–	uccidere	6	0.018

Within the PERLORIT list we found three verbs (two of which display two inflected forms each) that can be considered as adapted borrowings, i.e., istanziare/istanziamo (~ *to instantiate*, referring to manipulation of e.g., a variable), cicliamo/cicla (~ *to loop*, a method for repeating certain commands) and splittare (~ *split*, a command to segment text). Concordance analysis reveals that all of these borrowings are technical terms whose root derives from English. Let us take the example of splittare ~ *split*. As can be seen in Figure 5, in one case the term is paraphrased (Figure 5, line 3: "*… we have seen how to 'split,' i.e., how to divide a string into an array*"; translated by the authors), while in the two other cases it is used as a regular term, without further explanations.

Figure 5

Concordance lines of the verb splittare in Italian originals

```
1 (/pattern/, expression);  Nel caso più tipico, si può voler splittare una linea in parole, come segue: @words = split(/s
2 ndo scalare che specifica il numero massimo di campi in cui splittare il pattern. @array = split(/pattern/, expression);
3 s/guide_tutorials/Perl/Split_in_perl/ ) abbiamo visto come "splittare" cioè dividere una stringa in in un array.  Oggi
```

We found no occurrence of splittare (nor of its inflected forms) in the PERLTRIT corpus: here the English verb *split* is always translated with an Italian "standard" equivalent, e.g., dividere or spezzare.

FIGURE 6

Parallel concordance lines showing occurrences of the verb *split* in original English texts and their translations (selected)

```
1 | How can I split a [character] delimited string except when inside [character]?
  | -->perltrit: Come posso dividere una stringa delimitata da [carattere], tranne quando mi trovo tra [carattere]?
2 | You can't use "split (/,/)" because you shouldn't split if the comma is inside quotes.
  | -->perltrit: Non potete utilizzare "split(/,/)", poiché non dovete dividere se la virgola si trova tra virgolette.
3 | Each time you read a line of input, split it into a country and a city, look up the list of cities already known to
  | -->perltrit: Ogni volta che si legge una linea dell'input, la si spezza ("split") in un paese e in una città, si
4 | The regex determines the character sequence that "string" is split with respect to.
  | -->perltrit: La regex determina la sequenza di caratteri per i quali la "stringa" viene divisa.
```

In general, in PERLTRIT we found no adapted borrowings similar to those found in PERLORIT: all English-derived verbs either appear in a single text (e.g., matchare), or are accepted, integrated borrowings, also recorded in Italian monolingual dictionaries, and used by lay people with a moderate expertise in IT and computing (e.g., inizializzare ~ *initialize*, formattare ~ *format*).

As regards semantic loans, again we found anglicisms to be more numerous in PERLORIT than in PERLTRIT. For two verbs, i.e., processare (~ *process*) and valorizzare (~ *value*), the acquisition of a new word meaning, calqued on that of their English cognate, is evidenced by unusual collocational patterns. For instance, processare can only mean "bring to trial, prosecute" in standard Italian, and can thus only refer to human beings. A search for the verb lemma processare preceded or followed by a noun in a span of up to 4 words reveals that in PERLORIT the verb is used with the same meaning as *process* in English, as confirmed by the fact that some of their collocates are translation equivalents of each other (e.g., riga ~ *line*, and file). Processare is far less frequent in the PERLTRIT subcorpus, while manipolare (~ *manipulate*), a native Italian equivalent for *process*, turns up among the key verbs found in the translated texts (with collocates like file and variabili ~ *variables*), suggesting that most translators prefer the native Italian option.

In the case of ritornare, the third instance of a semantic loan we identified in the non-translated texts, collocational as well as *colligational* patterns are distorted. In Table 8 we show the noun collocates of transitive *return* in the English original subcorpus, and of ritornare in the Italian subcorpora.

TABLE 8

Frequency and top noun collocates of transitive *return*/ritornare in the PERL corpus

	PERLOREN	Fq	Fq / 1,000 words	PERLORIT	Fq	Fq / 1,000 words	PERLTRIT	Fq	Fq / 1,000 words
Verb (transitive)	return	628	2.104	ritornare	69	0.225	ritornare	12	0.037
Noun collocates (fq≥2)	value	106	0.355	valore	11	0.036	thread	2	0.006
	list	38	0.127	array	6	0.019			
	number	37	0.127	numero	6	0.019			
	values	25	0.083	valori	4	0.013			
	reference	18	0.060	reference	3	0.009			

As can be noticed, *return* and ritornare in the PERLOREN and PERLORIT subcorpora share as many as 4 noun collocates among the top 5 while PERLTRIT does not, thus suggesting that Italian authors tend to calque the English verb's collocational patterns to a greater extent than translators. Furthermore, the verb ritor-

nare, which means "go back to," can only be used intransitively in standard Italian, colligating with prepositions (meaning *from* and *to*) and with punctuation. These are the normal patterns observed in PERLTRIT, where only 12 occurrences of ritornare are followed by a direct object (vs. 69 that are used intransitively). In place of transitive ritornare translators opt for more standard Italian equivalents, e.g., restituire, which is the 3rd most typical verb of PERLTRIT (cf. Table 3). On the contrary, prepositions and punctuation hardly ever follow ritornare in PERLORIT, where the verb is mainly used transitively with the same meaning as *return* (a value, a list, a number etc.).

We observed a single case of a semantic loan preferred to a more standard Italian option by translators, i.e., uccidere (~ *kill*). While the standard Italian verb can only mean "cause the death of" and take animate objects, in PERLTRIT it collocates e.g., with processo (~ *process*), following the collocational pattern of English *kill* (*a process, the signal*); no occurrence of the verb uccidere was found in PERLORIT. A native equivalent would be terminare or bloccare, which can be found in the translated subcorpus to mean "stop (a process)," but with lower frequencies. These equivalents are also found in the non-translated subcorpus, but they do not make it to the top of the list of its most typical verbs, thus perhaps indicating that the whole notion of "stopping a process" is not so salient in the non-translated texts.

3.4.3. Morphosyntactic calques: -s plurals

An initial analysis of the words ending in *-s* in the Italian PERL subcorpora would lead us to think that the use of morphosyntactic calques is more frequent in translated than in original texts: PERLTRIT contains 144 word types of this kind with frequency ≥2 (1000 tokens), while PERLORIT contains 95 types (and 711 tokens). However, after careful manual inspection of concordances, aiming at excluding all the cases in which word forms are predominantly used outside actual Italian discourse (e.g., in code text, in untranslated English examples), the frequencies are reversed: original Italian texts contain slightly more calque types and substantially more calque tokens; see Table 9 for the complete results (after manual inspection and removal of calques found in the work of any single author).[18]

Table 9

Morphosyntactic calques in Italian original and translated texts

PERLORIT			**PERLTRIT**		
word	**fq**	**fq / 1,000 words**	**word**	**fq**	**fq / 1,000 words**
files	16	0.052	backticks	2	0.006
scripts	4	0.013	closures	2	0.006
expressions	3	0.009			

If we take as an example the word *files* in PERLORIT, concordances reveal that Italian writers use it within code text but also in running text (Figure 7). In translated texts, the word files, besides being less frequent (16 vs. 8 examples), either appears within code examples or within quoted English text (Figure 8, lines 1 and 2).

FIGURE 7

Concordance lines of files in original Italian texts (selected)

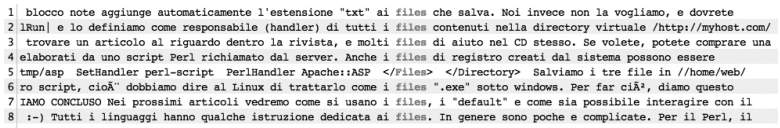
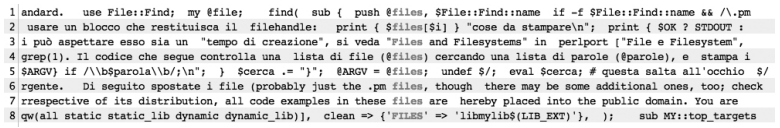


FIGURE 8

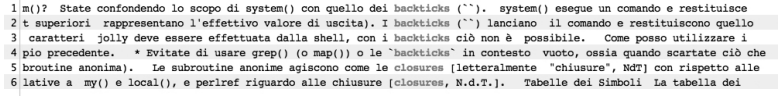
Concordance lines of files in translated Italian texts (selected)



In translated Italian texts, only two calques as defined above were found, i.e., backticks and closures. Concordances reveal, however, that these words are never used heedlessly by translators, who seem to be well aware of their status as foreignisms: their use is almost always accompanied by linguistic devices meant to signal their "otherness," such as exemplifications (Figure 9, lines 1 and 2), distancing devices like the use of inverted commas (Figure 9, line 4), and paraphrases (Figure 9, lines 5 and 6).

FIGURE 9

Concordance lines of backticks and closures in translations



3.5. Discussion

Taken together, the results presented above suggest that Italian translators of programming documentation seem less comfortable with anglicisms than technical writers of comparable texts: compared to the former, the latter use more unadapted borrowings when alternative Italian renderings are available; invent more verbs based on English roots and more homonyms of existing Italian verbs with new senses and new syntactic structures; preserve foreign morphological marks (*-s* plurals) to a greater extent.

While these three case studies make no claim to exhaustive coverage of anglicisms, the fact that results consistently point in the same direction does seem to suggest a trend toward normalization in this translation setting, or that the "law of growing standardization" predominates over the "law of interference," to use Toury's (1995) terms. One could argue that this translation setting is a marginal one: based on interaction with them and cursory checks of their profiles, the volunteer translators confirmed our expectations of being computer experts and to have had no formal education as language professionals. Therefore generalizations to professional translation settings should be done with caution. Yet we find that these results are all the

more surprising since they cannot be explained away in terms of training-induced prescriptivism, and could instead be viewed as candidates for task-inherent features of translation. The next and final section discusses implications of these results and of the study as a whole at several integrated levels.

4. Closing the circle: implications for CBTS

In this study we have presented a corpus-based analysis of anglicisms in original and translated Italian from the computing field, whose original inspiration came from classroom discussion on the legitimacy of borrowing as a translation strategy: to what extent is it appropriate to use English words when translating technical documentation in the programming field? Differences were observed in the behavior of writers and translators belonging the same discourse community and having similar educational and professional profiles, that pointed to a more conservative attitude on the part of translators. A careful reading of parallel concordance lines and several searches for alternative Italian renderings (in the two Italian subcorpora, in dictionaries, on the web) confirmed that differences could not be explained away in terms of unrelated variables (e.g., topic), and that they are indeed likely to be due to the translation process.

In closing we would like go back to Holmes' (1972) map of TS and reflect briefly on the relevance of our observations for descriptive, theoretical and applied branches of the discipline, and close, as Holmes did, on a meta-theoretical note. From the descriptive point of view, our case studies suggest that, if we take non-translated language as our baseline, translators in this well-defined discourse community normalize more than they transfer from English. Since translators and writers have very similar profiles (both groups are Perl programmers, and in fact some of the writers are also translators), we conclude that the differences observed might be due to a task effect, i.e., that the very act of translating may induce one to take a more conservative, normalizing attitude.[19]

This is just a tentative hypothesis, but agrees with the findings of Laviosa (2006: 272), who comes to similar conclusions in her study of anglicisms in business Italian, i.e., that "in translational language there seems to be a preference for native equivalents."

In terms of the theoretical debate over translation norms/universals, we believe that CBTS can contribute through the bottom-up accumulation of evidence about features of translation, obtained through small but carefully thought-out and clearly delimited studies, tweaking parameters so as to observe translator behavior under different conditions (e.g., varying the domain, the translator's profile/expertise/motivation, the target language and so on). Through the progressive exclusion of all other variables, it would thus be possible to isolate "phenomena […] for which it makes sense to produce a cognitively based explanation" (Malmkjær 2005: 18), thus ultimately confirming or refuting top-down hypotheses about the existence of translation universals. The process of excluding variables can be time-consuming and require some ingenuity, as shown by our case studies. Yet it is a necessary step: if we had relied on quantitative data about differences observed at the monolingual comparable level only, we would have been misled into coming to the wrong conclusion, i.e., that translated texts display *more* instances of interference than non-translated texts. Instead, we have shown that insights and hypotheses should emerge from the

accumulation of results of painstaking analyses conducted on closely comparable corpora, checked against their parallel text component(s) and/or taking into account alternatives offered by the target language.

Moving from theory to practice, our study argues against the *a priori* exclusion of translated texts from target reference corpora assembled for a translation task in a professional or didactic setting. It is generally assumed that translations will give "a distorted picture of the language they represent" (Teubert 1996: 247), and that translators should collect "original (not translated) examples of the types of TL texts they may be required to produce" (Vienne 1998: 114). Yet our case studies suggest that translators (even "amateur" ones) might in fact produce more carefully edited texts than writers (this will of course vary from setting to setting), such that their inclusion in a target language corpus alongside untranslated texts could be, under specific circumstances, appropriate.

Finally, and metatheoretically, we believe that the happy circularity we aimed for in this study (from applied concerns, through descriptive observations of theoretical and methodological relevance, back to applied recommendations), however difficult to achieve, is a valuable aspect of our discipline that could help bridge the chasm between theory and practice.

ACKNOWLEDGEMENTS

The authors would like to thank the issue editors, two anonymous reviewers and Federico Gaspari for insightful comments and careful reading of the manuscript. The research was partly funded by the University of Bologna under the "Progetti Strategici d'Ateneo" funding scheme.

NOTES

1. In this article we do not take sides in the debate about the theoretical tenability of the notion of "universal" (see e.g., the discussions in Chesterman 2004, Halverson 2003, House 2008, Toury 1995). We use and understand the term to refer to linguistic patterns repeatedly observed in translated texts that could point at underlying regularities of behavior (procedural or strategic) whose status (social, cognitive etc.) cannot be settled on the sole basis of corpus evidence.
2. *The size of the language industry in the EU.* Available at: <http://ec.europa.eu/dgs/translation/publications/studies/size_of_language_industry_en.pdf>, visited on 1 October, 2010.
3. "Practical Extraction and Report Language." See <http://www.perl.org/>, visited on 1 October, 2010.
4. The complete, original English text can be found in html format here: <http://perldoc.perl.org/perlretut.html>, visited on 1 October, 2010.
5. <http://pod2it.sourceforge.net/>, visited on 1 October, 2010.
6. <http://perl.enstimac.fr/>, visited on 1 October, 2010.
7. Several terms have been used in the literature, both within and outside of the corpus paradigm, to refer to the observed tendency for translated texts to be more normal or conservative (Baker 1996), conventional (Mauranen 2008), standard (Toury 1995), domesticated (Venuti 1995) or sanitized (Kenny 2001) with respect to their source texts and/or to comparable originals. We use "normalization" as an umbrella term for these various notions.
8. Ideally, one would also need a working definition of anglicisms and an objective way of classifying them (e.g., telling apart more or less integrated borrowings). Yet this is a theoretically complex question (see e.g., the discussion in Görlach 2003) that cannot be resolved here, and that would add little to this study, whose aim is not to shed new light on the notion of anglicisms, but rather to draw conclusions about the translation process based on the relative frequencies of anglicisms (however defined) in translated vs. non-translated texts.
9. <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>, visited on 1 October, 2010. Lemmatization for the Italian component was performed using Morph-It! (Zanchetta and Baroni 2005).

10. <http://cwb.sourceforge.net/>, visited on 1 October, 2010.
11. Readers interested in obtaining the corpus can contact the authors.
12. Because of the rich inflectional paradigm of the Italian verbal system, verbs borrowed from English tend to be naturalized through the addition of native suffixes. This is not the case for other parts of speech (e.g., nouns, adjectives) which therefore preserve their alien look.
13. The overwhelming majority of Italian words end with a vowel; according to the general rule the plural is formed by substituting the last letter of a word with either *i* (for masculine gender) or *e* (for feminine gender). There are very few native words ending with *-s* (even though of course there can be borrowings from languages other than English, or singular English words ending in *-s*, already covered by the first case study). For English words used in the plural in Italian the base form is normally used, i.e., with no (Italian or English) plural marks (cf. <http://www.accademiadellacrusca.it/faq/faq_risp.php?id=3781&ctg_id=93>, visited on 21 December, 2010).
14. <http://it.wikipedia.org/wiki/Pipe>, visited on 1 October, 2010.
15. Since *match* (from the translated keyword list) and *matching* (from the original keyword list) are identified as key for morphological reasons (translators use the base form, Italian writers the *–ing* form), their keyword status is in fact uncertain; therefore both words are disregarded here (even though the fact that writers opt for a more obviously foreign-looking word form might be relevant for our discussion).
16. The word *Hello* might be surprising in this context. The word is used in the sentence *Hello world/* Ciao mondo, which is often employed to exemplify basic scripts in Perl documentation.
17. Since these verbs are not part of the "common" Italian lexicon, they are unknown to the lemmatizer used for annotating the corpus, and thus cannot be "grouped" under the same lemma.
18. Six calques found in PERLORIT did not make their way to Table 9 due to the fact that each of them appears in the writing of a single author. These are *subroutines* (10), *backquotes* (6), *links* (4), *forms* (4), *cookies* (3) and *references* (2). While each choice is likely to be idiosyncratic, as a group they would seem to confirm the greater tolerance of anglicisms displayed by authors versus translators.
19. As suggested by an anonymous reviewer, one might argue that in an English-imbued work environment like that of computer programming, authors writing in Italian are in fact "thinking in English" and performing some form of mental, "inter-systemic" translation which might be akin to traditional, "inter-textual" translation. If this were the case, we would be comparing two forms of translation. While the hypothesis that translation and L2-influenced writing could be cognitively similar activities is an interesting one (suggestions in this area have been made by e.g., Cardinaletti 2005 and House 2004), our data and results point to differences rather than similarities: translation "proper" appears to trigger a more conservative behavior than (L2-influenced) writing. A different research setup would be required to shed light on this fascinating issue.

REFERENCES

Baker, Mona (1993): Corpus linguistics and translation studies: Implications and applications. *In*: Mona Baker, Gill Francis and Elena Tognini-Bonelli, eds. *Text and Technology*. Amsterdam: Benjamins, 223-250.

Baker, Mona (1995): Corpora in translation studies. An overview and suggestions for future research. *Target*. 7(2):223-243.

Baker, Mona (1996): Corpus-based translation studies. The challenges that lie ahead. *In*: Harold Somers, ed. *Terminology, LSP and Translation*. Amsterdam: Benjamins, 175-186.

Beeby, Allison, Rodríguez Inés, Patricia and Sánchez-Gijón, Pilar, eds. (2009): *Corpus Use and Translating*. Amsterdam: Benjamins.

Bernardini, Silvia and Zanettin, Federico (2004): When is a universal not a universal? Some limits of current corpus-based methodologies for the investigation of translation universals. *In*: Anna Mauranen and Pekka Kujamäk, eds. *Translation Universals. Do They Exist?* Amsterdam: Benjamins, 51-62.

Cardinaletti, Anna (2005): La traduzione: un caso di attrito linguistico. *In*: Anna Cardinaletti and Giuliana Garzone, eds. *L'italiano delle traduzioni*. Milano: Franco Angeli, 59-83.

Chesterman, Andrew (2004): Hypotheses about translation universals. *In*: Gyde Hansen, Kirsten Malmkjær and Daniel Gile , eds. *Claims, Changes and Challenges in Translation Studies*. Amsterdam: Benjamins, 1-13.

Englund Dimitrova, Birgitta (2004): Orality, literacy, reproduction of discourse and the translation of dialect. *In*: Irmeli Helin, ed. *Dialektübersetzung und Dialekte in Multimedia*. Frankfurt am Main: Peter Lang, 121-139.

Epstein, Brett Jocelyn (2010): Manipulating the next generation: translating culture for children. *Papers: Explorations into Children's Literature*. 20(1):41-76.

Furiassi, Cristiano and Hofland, Knut (2007): The Retrieval of False Anglicisms in Newspaper Texts. *In*: Roberta Facchinetti, ed. *Corpus Linguistics 25 Years On*. Amsterdam and New York: Rodopi, 347-363.

Görlach, Manfred (2003): *English Words Abroad*. Amsterdam: Benjamins.

Gottlieb, Henrik (2004): Danish echoes of English. *Nordic Journal of English Studies*. 3(2):39-65.

Halverson, Sandra (2003): The cognitive basis of translation universals. *Target*. 15(2):197-241.

Holmes, James S. (1972): The Name and Nature of Translation Studies. *In*: James S. Holmes, ed. *Translated! Papers on Literary Translation and Translation Studies*. Amsterdam: Rodopi, 67–80.

House, Juliane (2004): English as a *lingua franca* and its influence on texts in other European languages. *In*: Giuliana Garzone and Anna Cardinaletti, eds. *Lingua, mediazione linguistica e interferenza*. Milano: Franco Angeli, 21-48.

House, Juliane (2008): Beyond intervention: Universals in translation? *Trans-kom*. 1(1):6-19.

Johns, Tim (1991): Should you be persuaded – Two examples of data-driven learning materials. *English Language Research Journal*. 4:1-16.

Kenny, Dorothy (2001): *Lexis and creativity in translation: a corpus-based study*. Manchester: St. Jerome.

Kilgarriff, Adam (2001): Comparing Corpora. *International Journal of Corpus Linguistics*. 6(1):1-37.

Koby, Geoffrey S. and Baer, Brian James (2003): Task-based instruction and the new technology. Training translators for the modern language industry. *In*: Brian James Baer and Geoffrey S. Koby, eds. *Beyond the ivory tower: Rethinking translation pedagogy*. Amsterdam and Philadelphia: Benjamins, 211-227.

Laviosa, Sara (1998): The corpus-based approach: A new paradigm in translation studies. *Meta*. 43(4):474-479.

Laviosa, Sara (2002): *Corpus-based Translation studies: Theory, Findings, Applications*. Amsterdam and Atlanta: Rodopi.

Laviosa, Sara (2006): Data-driven learning for translating anglicisms in business communication. *IEEE transactions on professional communication*. 49(3):267-274.

Malmkjær, Kirsten (2005). Norms and nature in translation studies. *Synaps: Fagspråk, kommuniksjon, kulturkunnscap*. 16:13-19.

Mauranen, Anna (2008): Universal tendencies in translation. *In*: Gunilla M. Anderman and Margaret Rogers, eds. *Incorporating Corpora: The Linguist and the Translator*. Clevedon: Multilingual Matters, 32-48.

McEnery, Tony, Xiao, Richard and Tono Yukio (2006): *Corpus-based Language Studies: An Advanced Resource Book*. London and New York: Routledge.

Olohan, Maeve (2004): *Introducing corpora in translation studies*. London and New York: Routledge.

Olohan, Maeve (2007): Economic trends and developments in the translation industry. *The Interpreter and Translator Trainer*. 1(1):37-63.

Pearson, Jennifer (1998): *Terms in Context*. Amsterdam and Philadelphia: Benjamins.

Piqué-Angordans, Jordi, Posteguillo, Santiago and Melcion, Lourdes (2006): The development of a computer science dictionary, or how to help translate the untranslatable. *In*: Elisabet Arnó Macià, Antonia Soler Cervera and Carmen Rueda Ramos, eds. *Information Technology in Languages for Specific Purposes. Issues and Prospects*. New York: Springer, 213-229.

Rayson, Paul and Garside, Roger (2000): Comparing corpora using frequency profiling. *Proceedings of the ACL Workshop on Comparing Corpora*. 38th annual meeting of the Association for Computational Linguistics (1-8 October 2000, Hong Kong). 1-6.

Teich, Elke (2003): *Cross-linguistic Variation in System and Text*. Berlin and New York: Mouton de Gruyter.

Teubert, Wolfgang (1996): Comparable or parallel corpora? *International Journal* of *Lexicography*. 9(3):238-264.

Toury, Gideon (1980): *In Search of a Theory of Translation*. Tel Aviv: The Porter Institute for Poetics and Semiotics.

Toury, Gideon (1995): *Descriptive Translation Studies and Beyond*. Amsterdam and Philadelphia: Benjamins.

Venuti, Lawrence (1995): *The Translator's Invisibility: A History of Translation*. London and New York: Routledge.

Vienne, Jean (1998): Teaching what they didn't learn as language students. *In*: Kirsten Malmkjær, ed. *Translation and Language Teaching. Language Teaching and Translation*. Manchester: St. Jerome, 111-116.

Williams, Donna (2005): *Recurrent Features of Translation in Canada: A Corpus-based Study*. Doctoral thesis, unpublished. Ottawa: University of Ottawa.

Zanchetta, Eros and Baroni, Marco (2005): Morph-it! A free corpus-based morphological resource for the Italian language. *Proceedings of Corpus Linguistics 2005*. Birmingham: University of Birmingham. 1(1).