Human–computer Interaction in Pun Translation

Waltraud Kolb
https://orcid.org/0000-0002-7806-6478

Tristan Miller
https://orcid.org/0000-0002-0749-1100

Abstract

Puns and wordplay are stylistic features that are recurrent in many literary and creative text types but are among the most challenging for humans to translate. At the same time, they have proved impervious to traditional machine translation. In this chapter, we present and evaluate PunCAT, an interactive electronic tool for the translation of puns, designed to provide specialized support to human translation workflows. Our evaluation is based on an empirical pilot study in which nine graduate students translated six English puns taken from literary works and films into German, with and without PunCAT. Combining computational-linguistic and cognitive approaches, we triangulated logging data from PunCAT and the keylogger Inputlog, verbal data from questionnaires, handwritten notes, and annotations, as well as target texts. Fine-grained analyses of the participants' translation and decision-making processes and their interaction with the tool show that PunCAT effectively supports the translation process in terms of stimulating brainstorming and broadening the translator's pool of solution candidates, and we have also identified a number of directions in which the tool could be adapted in the future to better suit translators' work processes.

Introduction

Wordplay causes tremendous difficulties for translators, and so is a widely studied phenomenon in the field of translation studies. Despite this, and the trend in recent years to technologize the translation process, little attention has been paid to the use of computers for the translation of wordplay. This is because most language technology, including machine translation (MT), has been developed for use with informational rather than literary and other creative texts. As such, existing digital tools and resources tend to ignore linguistic anomalies and ambiguities, or else to treat them as imperfections to be eliminated rather than preserved.

Punning is a ubiquitous form of wordplay in which one word is used to evoke another word with a similar or identical pronunciation. Puns pose special challenges over many other types of wordplay in that they rely not just on surface-level features, but also a relatively sophisticated understanding of lexical semantics and (usually) the complex pragmatic phenomenon of humour. It is for these reasons that puns are often held to be untranslatable; while this view is overly pessimistic with respect to human translation, it is true that puns are impervious to general-purpose MT. Nevertheless, recent advances in computational semantics have brought us to the point where language technology might now play a useful role in the translation of puns by providing specialized support to existing translation workflows. Of course, the idea of

computer-mediated translation is by no means a new one (see Kay 1980), though the present chapter is the first time, to our knowledge, that it has been empirically explored with respect to wordplay, bringing together computational-linguistic and cognitive approaches.

This chapter presents and evaluates PunCAT, an interactive electronic tool for the translation of puns. Following the strategies known to be applied in pun translation (Delabastita 1996; Low 2011), PunCAT automatically translates each sense of the pun separately; it then allows the user to explore the semantic fields of these translations in order to help construct a plausible target-language solution that maximizes the semantic correspondence to the original. Our evaluation is based on an empirical pilot study in which the participants translated English puns into German, with and without PunCAT. We aimed to answer the following questions: Does the tool support, improve, or constrain the translation process? If so, in what ways? What are the tool's main benefits as perceived and described by the participants? Our analysis of the translators' cognitive processes gives us insight into their decision-making strategies and how they interacted with the tool. We find clear evidence that PunCAT effectively supports the translation process in terms of stimulating brainstorming and broadening the translator's pool of solution candidates. We have also identified a number of directions in which the tool could be adapted to better suit translators' work processes.

Background

Punning is a rhetorical device in which one word or phrase is used to evoke the meaning of another word or phrase with the same or slightly different pronunciation. The term *pun* can refer to the complete expression containing this ambiguity, or more specifically to the particular word or phrase that carries the double meaning; the *target* is the latent word or phrase that is evoked. (This linguistic sense of *target* should not be confused with the translatological sense—i.e., the text that a translator produces.) For example, in the text, "The sign at the nudist camp read, 'Clothed until April'," the pun "clothed" evokes the target "closed".

Though the phonological and semantic mechanisms behind punning are well understood theoretically (Hempelmann and Miller 2017), the translation of puns is not straightforward (Delabastita 1996; Vandaele 2011; Knospe, Onysko, and Goth 2016). Puns are frequently what Angelone (2010, 18), in the context of cognition and uncertainty management in translation, has described as a problem nexus ("the confluence of a given textual property and level. . . intersecting with some sort of deficit in the translator's cognitive resources") where the natural flow of translation is interrupted or impeded. Nevertheless, there are a number of high-level strategies that translators have at their disposal. Delabastita (1996) presents a typology of eight methods, including PUN→PUN (replacing the source-language pun with a target language pun, possibly with different semantics, structure, or function), PUN→NON-PUN (substituting non-punning language that preserves one or both of the meanings), PUN→PUNOID (replacing the pun with some non-punning wordplay or rhetorical device), and PUN→ZERO (omitting altogether the language containing the pun). The choice of strategy for any given pun depends on various factors (Klitgård 2018), and while strategies that preserve wordplay are generally preferable, they are often the most challenging to pull off.

It is therefore reasonable to ask whether computers might play some role in the translation of puns. Indeed, researchers have been taking an increasing interest in the use of language technology in creative translation in general, including the integration of MT systems into human translation workflows (Moorkens et al. 2018; Toral and Way 2018; Taivalkoski-Shilov 2019;

Jiménez-Crespo 2020). However, since punning works by subverting linguistic conventions, puns are not suitable for off-the-shelf, end-to-end MT systems, particularly those based on the prevailing neural paradigm (Miller 2019). And while others have pointed out the potentials of digital tools to assist literary translation processes (Youdale 2019), no currently available tool specifically supports the translation of puns.

That said, there does exist a body of work in natural language processing (NLP), including some specifically concerned with puns, that could be leveraged to develop such a tool. This includes computational models of the phonological properties of puns; algorithms to determine whether a given passage contains a pun and if so, to pinpoint its location; and approaches for automatically interpreting puns by recovering the target word and identifying the double meaning (by referring to word senses listed in a given dictionary); as well as various more fundamental, general-purpose methods and resources, such as grapheme-to-phoneme models, multilingual semantic networks, and methods for measuring the semantic or phonetic similarity between words. Miller (2019) briefly surveys most of this work and outlines how it could be synthesized into a computer-assisted translation (CAT) tool for puns. Taking inspiration from Kay (1980), and consistent with a functional approach to translation (Reiß and Vermeer 1984; Nord 2018), the ultimate aim of such a tool would be to help translators produce a viable target text that performs its function in the target situation (i.e., creating a humorous effect), preferably without losing the wordplay.

Experiment

As we have shown, there exists a considerable amount of past work on puns in the fields of linguistics, NLP, and translation studies, as well as a proposal for how these hitherto separate channels of research might be applied to the construction of a real-world tool to support pun translation. In this section, we describe a prototype of such a tool and evaluate its usefulness in a user study.

PunCAT

PunCAT is the second author's partial implementation of the CAT tool proposed in Miller (2019).[1] As originally envisaged, the tool was to scan a complete source document to automatically locate all puns and then to interpret them, both lexically and semantically, by identifying the respective words and meanings with reference to an electronic dictionary. The tool would then present each interpreted pun in context, along with literal translations of the two meanings, and then allow the user to interactively explore the lexical-semantic space to find pairs of words that might form similar or equivalent puns in the target language. Since our interest lies in how human translators interact with such a tool to produce translations, rather than in the accuracy of the pun detection and interpretation algorithms, in PunCAT we chose to fully implement only those parts of the tool concerned with finding translation candidates. While PunCAT's user interface (UI) does present the user with a list of interpreted puns to translate, we performed the selection and interpretation of these puns ourselves and hard-coded this data into the program's input files; the integration of fully automatic methods for these tasks (Miller, Hempelmann, and Gurevych 2017) is left for future development.

PunCAT's UI is illustrated in Figure 1, with four distinct regions of functionality labelled A through D:

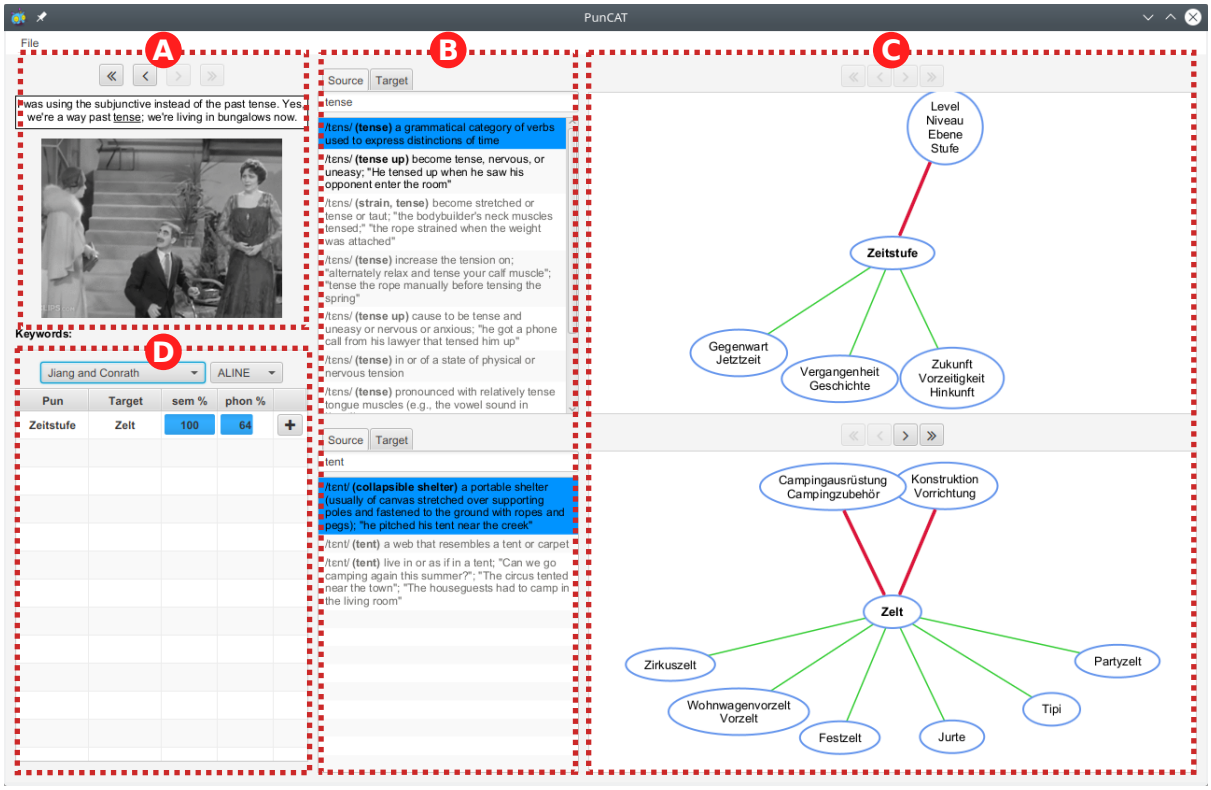

Figure 1: PunCAT's UI, showing (A) the source pun in context, (B) the words used in the pun and their meanings, (C) the lexical-semantic graphs of these meanings in the target language, and (D) the translation candidate list.

A. The source material shows the pun in its original textual context, along with an associated illustration or movie clip in the case of (audio-)visual media. The word being punned upon is automatically underlined by the system. Above the text are navigation buttons for browsing forward and backward through the list of source texts.

B. The dictionary contains "Source" and "Target" tabs. The former shows, at least initially, the two words used in the source-text pun along with the respective dictionary definitions for each of their possible senses; the two senses used in the pun are automatically highlighted by the system. The "Target" tab (not activated in Figure 1) has a similar appearance, except that it shows words and dictionary definitions in the target language. It is initially populated with the two words that are the direct translations of the selected senses of the two source-language words. Users can look up other source- or target-language words by typing them into the respective text fields, and can select any of those words' senses. When a user selects a new sense in the "Source" tab, the system automatically populates the corresponding word and sense in the "Target" tab.

C. The lexical-semantic network shows two graphs of concepts in the target language. Each graph node corresponds to a concept, and shows a list of synonymous target-language words for that concept. The nodes at the centres of the two graphs correspond to the senses currently selected in the two "Target" tabs of the dictionary. A concept may have more than one synonym; the currently selected one is printed in boldface. The nodes are arranged into a tree structure, with hypernyms (more general concepts) of the central node connected with thick red lines and hyponyms (more specific concepts) with thin green

lines. To avoid overcrowding the graph display, the system puts a limit on the number of hypernym and hyponym nodes displayed at once, but users can bring others into view by using the navigation buttons at the top of each graph. Users can also re-centre a graph by clicking on a word in one of the other nodes.

D. The candidate list always shows, at minimum, the two currently selected target-language words (labelled "Pun" and "Target"), along with a measure of the phonetic similarity between them ("phon %") and a measure of the semantic similarity to the source-language pun ("sem %"). (More specifically, similarity is measured by calculating the semantic similarity between the sense that was initially selected in the upper graph—i.e., the sense that the system decided corresponded to one of the meanings of the source-language pun—and the sense that is currently selected in the upper graph, then adding to this the semantic similarity between the senses initially and currently selected in the lower graph, and then dividing the result by two.) By default, phonetic and semantic similarity are calculated using the ALINE (Kondrak 2000) and Jiang and Conrath (1997) metrics, respectively, though others may be selected. Whenever the user finds a pair of target-language words that they feel could form the basis of a punning joke, they can add them to the exportable candidate list.

PunCAT has been developed in a language-independent manner, and so should in theory work with any pair of source and target languages for which there exist electronic lexical-semantic networks, interlingual links between the concepts of the two networks, and pronunciation information in the form of electronic pronouncing dictionaries or grapheme-to-phoneme models. For our experiments, we use WordNet (Fellbaum 1998) and GermaNet (Hamp and Feldweg 1997; Henrich and Hinrichs 2010) as the English and German semantic networks, and grapheme-to-phoneme models produced using Phonetisaurus (Novak, Minematsu, and Hirose 2016).

It should be noted that although WordNet and GermaNet are among the largest lexical-semantic networks, they still have significant (and asymmetric) gaps in their coverage of words and concepts, and even among the concepts they share, many of the corresponding interlingual links are missing. If a PunCAT user enters a word in the "Source" tab that cannot be found in WordNet, the tool will inform the user of this and blank out the corresponding "Target" tab entry and semantic network. This will also happen if the user selects a sense in the "Source" tab that has no interlingual link to a GermaNet sense, or if the user enters a word in the "Target" tab that does not exist in GermaNet.

Source data

We aimed to test PunCAT with a variety of pun and source text types. In selecting our examples, our overriding concerns were that (a) there should exist at least one published translation for each of them to serve as a basis of comparison, and (b) at least half of these known translations should use the PUN→PUN strategy and the two senses of the target pun should exist in GermaNet. These conditions ensure that for at least half of our texts, the pun from a reference translation could in theory be "discovered" using PunCAT. We also imposed the requirement that, for all source-language puns, the two words and meanings exist in WordNet, so that PunCAT would at minimum provide the user with the complete and correct interpretation of the pun. However, for some of our examples, there was no concept in GermaNet corresponding to either or both of the WordNet concepts, or else these concepts did exist in GermaNet but were not linked to the corresponding WordNet ones. We admitted these examples in order to see whether PunCAT can still lead to a viable translation under less than ideal conditions.

#	Text	Pun gloss	Source
1	"And how many hours a day did you do lessons?" said Alice. . . "That's the reason they're called lessons," the Gryphon remarked: "because they lessen from day to day."	lessons: classes lessen(s): diminish(es)	*Alice's Adventures in Wonderland* (1865 novel by Lewis Carroll)
2	When they operated on him she prepared him for the operating table; and they had a joke about friend or enema.	enema: rectal injection enemy: adversary	"A Very Short Story" (1924 short story by Ernest Hemingway)
3	"Hold on, everyone. This ain't gonna be no picnic."	picnic: outdoor meal picnic: easy task	*Cloudy With a Chance of Meatballs 2* (2013 screenplay by Erica Rivinoja)
4	Nemo: "What's that?" Nemo spots a dive boat. . . Tad: "I know what that is. Sandy Plankton saw one. He said it was called. . . a butt!" Pearl: "Wow. That's a pretty big butt."	butt: hindquarters boat: seagoing vessel	*Finding Nemo* (2003 screenplay by Andrew Stanton, Bob Peterson, and David Reynolds)
5	"Then we play somewhere where the Guild won't find us," said Glod cheerfully. "We find a club somewhere—" "Got a club," said Lias proudly. "Got a nail in it." "I mean a night club," said Glod. "Still got a nail in it at night."	club: discotheque club: bludgeon	*Soul Music* (1994 novel by Terry Pratchett)
6	"I should say you are intruding! I should say you WERE intruding, pardon me. I was using the subjunctive instead of the past tense. Yes, we're a way past tense; we're living in bungalows now."	tense: verb form tents: portable shelters	*Animal Crackers* (1930 screenplay adapted by Morrie Ryskind)

Table 1: Puns used in our pilot study.

The six puns we selected, their glosses, and their immediate contexts, are shown in Table 1. For each of these contexts, we manually located the pun, identified the corresponding words and senses in WordNet, extracted any corresponding illustration or film clip from the source material, and then used this data to prepare an input file for PunCAT. Because these contexts are not long enough to allow for an accurate interpretation of the humour, we also prepared a hard copy that reproduced the six texts in a wider context—i.e., a few sentences before and/or after the text containing the pun, along with any corresponding film stills. The hard copy did not, however, mark up the location or meaning of the pun.

Experimental setup

The participants in this study were nine students from the Master's program in translation at the University of Vienna's Centre for Translation Studies. (These participants, coincidentally all women, had responded to a general call for volunteers sent to 170 students with English as one of their working languages.) The experiment took place at the Centre's media lab, which allowed us to exert a greater degree of control and oversight over the translation process, but also meant that the participants could not work in their usual, authentic working environments. Each participant was provided with a workstation on which PunCAT, Microsoft Word, a web browser, and a keylogger (Inputlog) were preinstalled.

The experiment consisted of two 45-minute sessions, during each of which the participants translated three of the puns from Table 1. In Session 1, they translated three puns without the help of PunCAT, and in Session 2 they translated three different puns with the help of PunCAT. In both sessions, participants were free to consult outside resources, including accessing the Internet through the preinstalled web browser. However, we requested that they not seek out existing translations of the source texts. The participants were divided into two groups: Group A translated Puns 1–3 without the tool and Puns 4–6 with it, while Group B translated Puns 4–6 without the tool and Puns 1–3 with it. This way, each participant worked in both modes (with and without PunCAT) and for each pun, target texts that had been produced with and without the tool were available for analysis. Participants were encouraged to produce target solutions consistent with the general function of the selected passages (i.e., creating a humorous effect). We indicated that producing target solutions that were themselves puns was preferable but not necessary.

Before the first session, we explained to participants the aims and setup of the study and presented Delabastita's (1996) typology of pun translation strategies. Having the participants work without the tool in Session 1 had the advantage of allowing them to get accustomed to pun translation without the additional pressure of learning to use any new technology. PunCAT was introduced before the start of Session 2, and participants had some time to test-drive it on an example pun (not one of the six from Table 1). Participants were aware that the second author, with whom they had no prior relationship, had implemented the system.

In both sessions, the participants wrote their translations in Word, with the keylogger recording all keyboard activity, mouse movements, and access to outside resources (Internet and PunCAT). PunCAT itself also logged all interactions, including the text of all "Source" and "Target" tab dictionary lookups and which concept nodes and words were hovered over or clicked on in the lexical-semantic networks. At the start of each session, participants were given their three source puns on hard copy as well as blank note paper; they were encouraged to make whatever annotations and notes they wished and asked to submit these at the end of the experiment. After Session 2, participants filled out a questionnaire (Q1) containing questions on their background

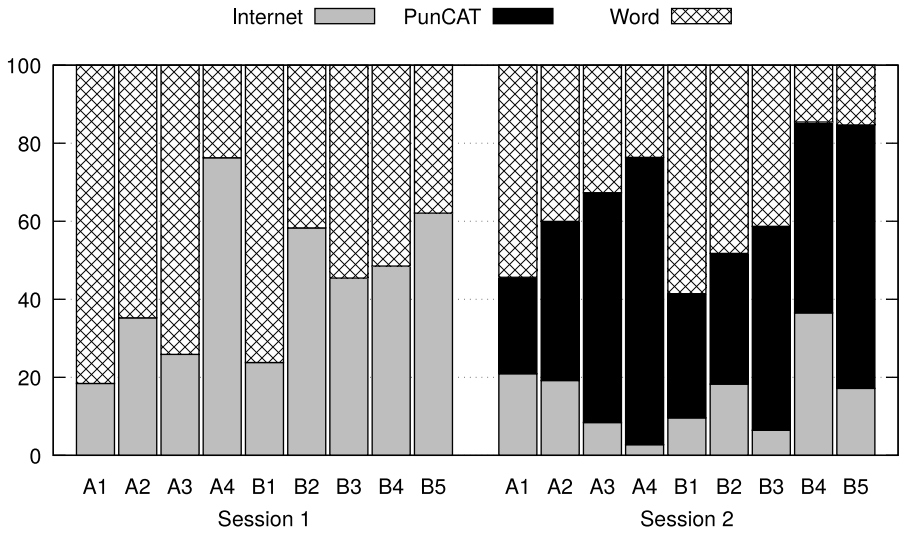

Figure 2: Percentage of participants' total work time by session and computer activity.

and their experiences working on the puns with and without PunCAT. Three days after the experiment, they answered a second questionnaire (Q2) in which they listed any further solutions that had occurred to them and commented on the published translations of the puns they had worked on, comparing these versions to their own.

Analysis and discussion

The triangulation of logging data from the keylogger and PunCAT, together with the answers to the two questionnaires and the handwritten notes and annotations, allows us to trace the participants' working processes in great detail. Given the aims of this study, the focus here will be on the participants' interactions with PunCAT and the role these interactions play in the context of the overall translation process. In this section, we will first give a brief overview of the logging data before discussing the participants' working strategies and decision-making processes in detail, bringing together temporal and cognitive aspects. As it will be impossible to discuss all 62 translations that were produced in the course of the experiment (some participants having produced more than one translation for some puns), nor all the published translations, we will refer to them only where relevant.

Our setup imposed certain restrictions, first and foremost that the participants worked at the university's premises and not in their usual working environments, and that each session was limited to 45 minutes. Participants were therefore not fully free to set their own pace of work, although within each session they were able to follow their own rhythms. Figure 2 shows how the participants distributed their working time. Nearly all participants spent less time on research during Session 1 than during Session 2, which is not surprising given that during Session 2 they had both PunCAT and the Internet at their disposal. Comparing Internet use only, eight out of nine participants spent (in some cases much) less time on the Internet during Session 2, when they also used PunCAT, than during Session 1, when the Internet was the only outside resource. However, both Internet and PunCAT interaction times vary greatly across participants. In the case of the Internet, interaction times ranged from 18.4% (A1) to 76.3% (A4) of the participants' total working times during Session 1, and from 2.7% (A4) to 36.5% (B4) during Session 2; in

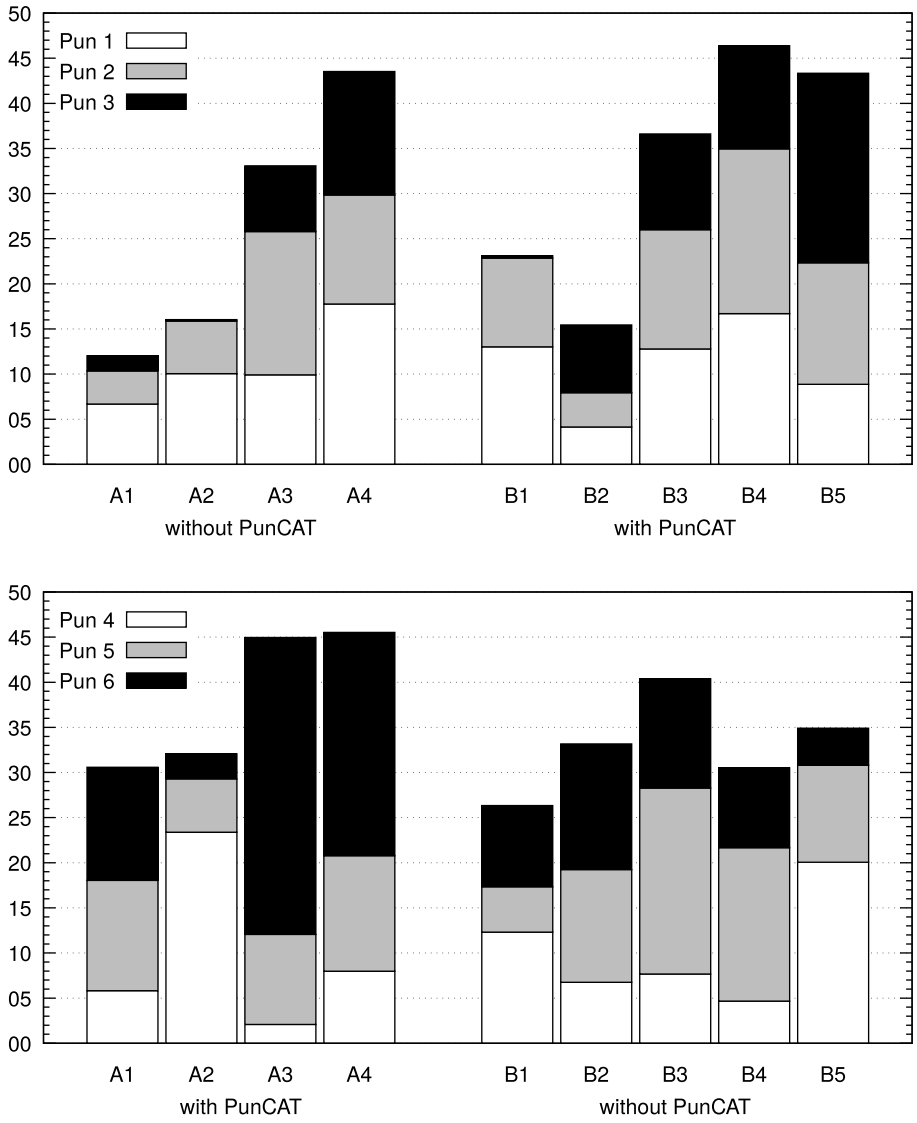

Figure 3: Participants' total work time per pun, in minutes.

the case of PunCAT, interaction times ranged from 24.7% (A1) to 73.7% (A4). This variation is an indication of very different working styles, something that will also emerge below.

Figure 3 shows the total working times participants spent on each pun, and Figure 4 the average work time across all puns translated with or without PunCAT. As previously mentioned, the passages given to the participants in hard copy included somewhat wider contexts, and the participants were told that they did not necessarily need to translate the full texts, but rather just the passages containing the puns. Though some translations include the whole text, Figures 3 and 4 report only those periods where the translator was working on the pun and its immediate context. Both figures are inconclusive in that there is no clear indication whether using PunCAT speeds or slows the translation process. Six out of nine participants were slower on average when they translated with the tool than without (Figure 4). However, this needs to be interpreted against a number of other factors: For one, the participants were not yet well acquainted with the tool (see below), and the exploration that it afforded might have led some to do more brainstorming than would have been absolutely necessary in order to come up with at least one

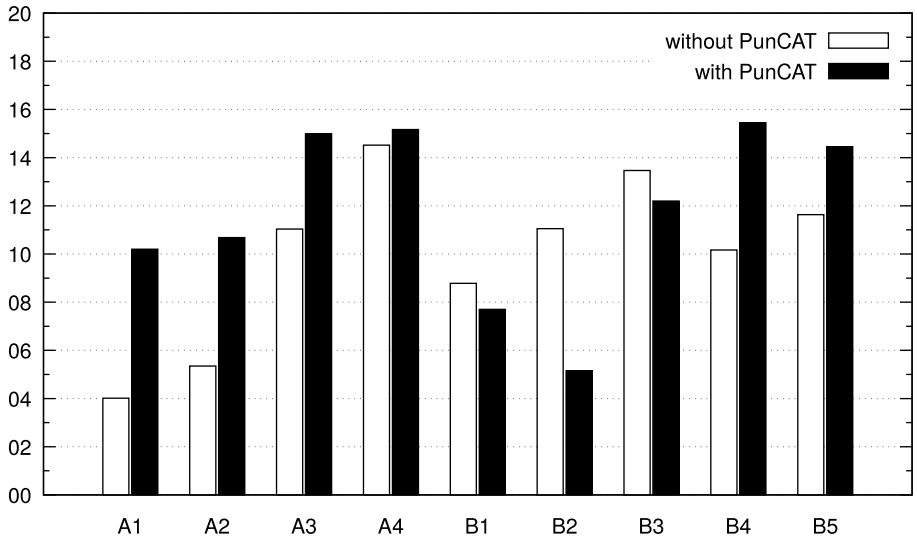

Figure 4: Participants' average work time per pun, in minutes.

solution. Another factor is that not all puns posed the same level of difficulty, and even the level of difficulty of a pun will differ from translator to translator. What the numbers and the high degree of intersubject variability do underline, though, is again the role that personal working styles and profiles play in the translation process.

Figure 5 illustrates the participants' interactions with PunCAT: the numbers of graph nodes they hovered over and clicked (the latter implying a somewhat deeper engagement with a particular term or concept) and the numbers of additional source or target language terms manually typed in. Again, we can see that these indicators of interaction with the tool vary greatly. While some participants hovered over very high numbers of nodes for a particular pun, others accessed only a handful of nodes throughout.

As this study's focus is on interaction with PunCAT, in the following sections we present some fine-grained observations and analyses of the translators' cognitive processes, working styles, and outputs, based on the logging data; recourse will be made to verbal data from the two questionnaires and handwritten notes where appropriate. First, we will look at how participants from Group A (who translated Puns 4, 5, and 6 with PunCAT) made use of the tool; second, we will explore Group B, who translated Puns 1, 2, and 3 with the tool.

Group A

Participants from Group A spent a greater proportion of Session 2 working with the tool than Group B did (49.5% vs. 46.8% on average; Figure 2). We will start our observations with participant A4, who interacted longer with PunCAT than anyone else (73.7% of Session 2). Her overall working style can be described as highly systematic—more so than that of her colleagues. (By comparison, the working processes of B4 and B5 were highly fragmented, with frequent switches between puns and between applications.) In both sessions, she first spent a long, contiguous period doing research and brainstorming, whether on the Internet (Session 1) or in PunCAT (Session 2). Then, once apparently satisfied with target candidates she had found, she proceeded to speedily type her translation; she made some local revisions while producing the target sentences, but did not switch much back and forth between the writing and research

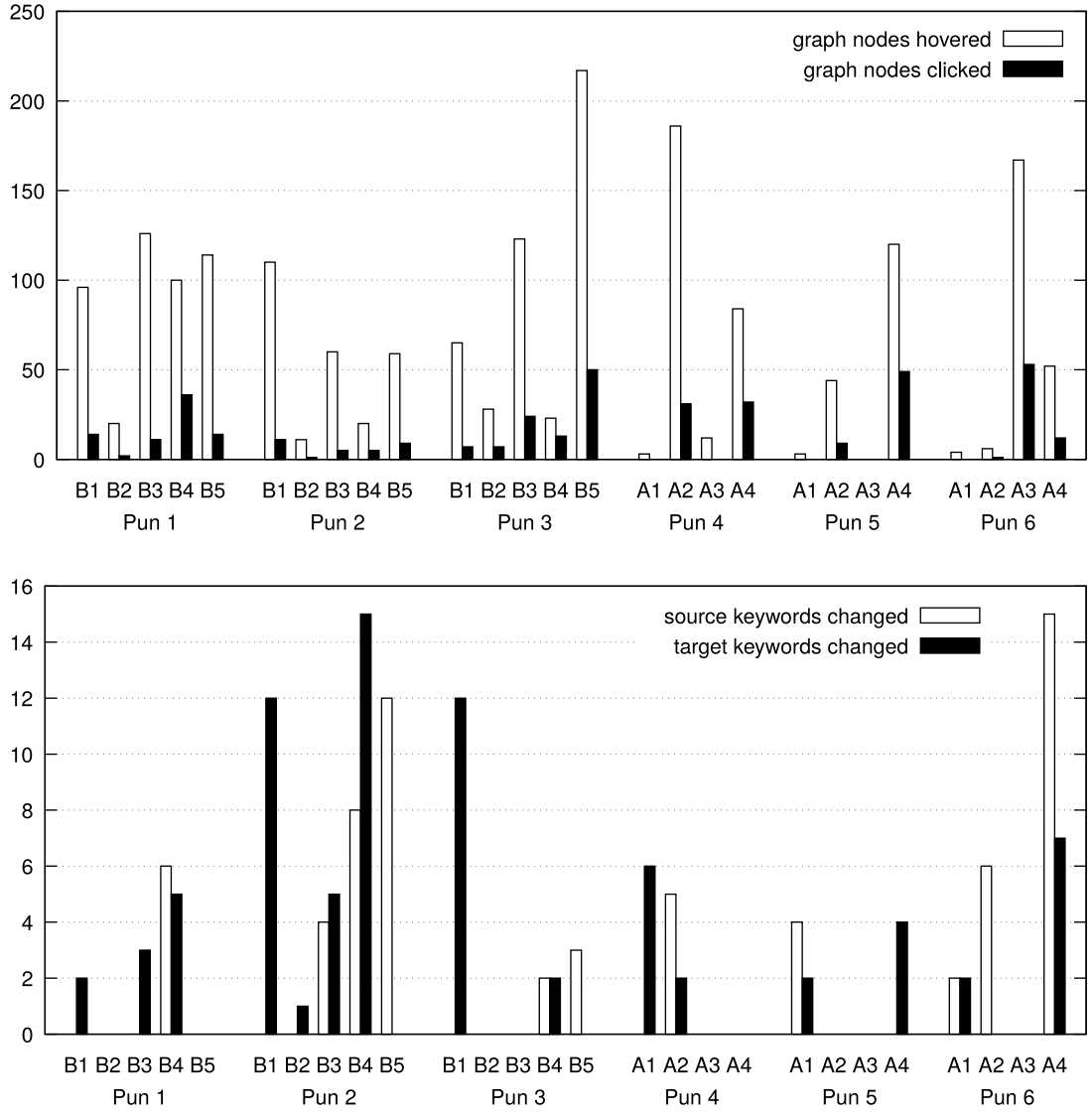

Figure 5: PunCAT interaction statistics by participant and pun.

modes. Her mind, it seems, was largely made up at the end of her research phase. The greatest part of her overall time went into research and brainstorming, while she spent relatively little time in Word. Our data do not always show exactly what a participant did at any given moment; for instance, we cannot precisely allocate periods of source- or target-text reading or general reflection, something that could be studied with eye-tracking software or think-aloud protocols. But our data do indicate that she did most of her brainstorming and deciding during her research phase; that this seems to be her preferred working style is also reflected by the fact that in both sessions she distributed her efforts almost identically between the applications (Word: 23.7% and 23.6%, respectively; research tools: 76.3% and 76.4%, respectively). During Session 1, she consulted a very large number of outside resources (96 task bar clicks) and spent more time than anyone else on the Internet; during Session 2, she spent more time than anyone else in PunCAT. What is particularly interesting is that with the exception of looking up the meaning or possibly spelling of a German term in an online dictionary while working on Pun 6, she did not consult any other outside resources during Session 2. As a consequence, out of the five target puns she produced in this session (including three versions for Pun 6), four were exclusively retrieved from or prompted by her search in PunCAT; only one was unrelated. (In Session 2, when the participants worked with PunCAT, all except B4 and B5 translated at least one pun without any Internet resources—i.e., relying exclusively on PunCAT and their own skills.) In all five cases, A4 adhered to the PUN→PUN strategy and came up with viable solutions, so it is worthwhile to take a closer look at how she made use of the tool.

In the case of Pun 4, the first one she translated with PunCAT, the tool offered a large number of names of different types of boat and different parts of the human body; all in all, she hovered over 84 graph nodes, clicked on 32 (Figure 5), and finally selected two pairs of words from this pool for her candidate list: *Arsch/Arche* and *Po/Boot*. Per her handwritten notes, the reason she eventually discarded the first pair ("arse"/"ark"), despite the good phonetic overlap, was that she judged *Arsch* too vulgar for the film's young target audience. *Po*, on the other hand, is an inoffensive colloquial term for "butt" (and *Boot* is "boat" in English). Her final version reads,

> Tad: „Er sagt, es heißt . . . Po. . . po. . . Boot!"
>
> Pearl: „Wow, das ist ein richtig großes Popoboot."

PunCAT had given her both *Po* and *Popo*, the reduplicative variant being very much part of the language of children. She creatively combined the two proposed words so that Tad's "*Po. . . po. . . Boot*" is as hesitant as it is in the original, and the logic of Pearl's answer arguably even more humorous than in English. The English terms are incidentally phonetically very similar to their direct German counterparts, which might be responsible for the fact that, with three exceptions, all nine participants worked with some variation of *Po*, as did the German translators of the distributed film.

When she worked on Pun 5, she looked at even more words offered by PunCAT for buildings and weapons (120 nodes hovered, 49 clicked; Figure 5) and added four target keywords of her own (also referring to buildings and weapons), thereby exploring the semantic fields of "club" used in the source text. PunCAT did provide her with two building/weapon pairs she liked well enough to enter into her list of candidates: *Schuppen/Knüppel* and *Kneipe/Peitsche*; in both cases the phonetic overlap is at best modest, which is probably what made her continue her search. The third candidate pair on her list, which she eventually used for her final target pun, contains perfect homonyms: *Schuppen/Schuppen*. The first of these is a colloquial term (in the singular) for a pub or a bar, and the second (plural) meaning "dandruff". The homonymy makes it a very

convincing solution that nicely fits the scene, though it does not fully preserve the semantics of the original pun. Just as with Pun 4, we can observe how a word proposed by PunCAT seems to have stimulated a burst of creativity: Although the "dandruff" sense of *Schuppen* is not given in GermaNet (and therefore in PunCAT), the "pub" sense is, and this is evidently what prompted her to come up with the missing homonym herself. A gloss of her translation would read,

> "We will look for a *Schuppen*$_1$ somewhere—"
>
> "I've got *Schuppen*$_2$," Lias said proudly. "Quite a lot, actually."
>
> "I meant a *Schuppen*, a building where we can play," Glod said.
>
> "I can play in a building also with my *Schuppen*$_2$."

While A4 spent nearly eight minutes on Pun 4 and nearly thirteen on Pun 5, she worked considerably longer on Pun 6 (almost 25 minutes; Figure 3). This might have to do with the fact that she had extra time on her hands completing the first two puns, and she did in the end produce three different target versions for Pun 6. Of the three target solutions, two were retrieved from the PunCAT pool, while she came up with the third herself. She did, however, subsequently check whether the tool contained the words she had used, and she found that it did. The version she came up with herself was the one she liked best and ranked first. This solution is again a homonym pair that works well in the context as it preserves the original's allusion to the passage of time, it reads (in back-translation), "I should have used the *Perfekt*. Yeah, nowadays everything needs to be *perfekt*," the first term being a grammatical term for the present perfect tense, and the second meaning "without defect".

The answers she gave in her Q1 indicate that working with the tool was a positive experience for her: She "agreed" that the tool was easy to use and "strongly agreed" that she found it useful and could imagine using it in the future. Also, her satisfaction with her target texts is somewhat higher for the PunCAT group of puns than for those she had translated without the tool. All in all, she found the tool "pleasant to use" and "quite intuitive". She also thought she had saved time using the tool "because no notes, no extra searching for synonyms etc." was necessary. In fact, the average time she spent per pun in Session 2 was slightly longer than in Session 1 (15m10s compared to 14m31s; Figure 4), but this must be interpreted in light of the fact that she produced three versions for Pun 6. If we disregard the time she worked on the two additional versions for Pun 6, the average would indeed be lower for Session 2 (12m49s), thus confirming her own intuitive impression (but, as pointed out above, no general conclusions can be drawn about time savings).

While A4 spent 73.7% of her total working time in Session 2 in PunCAT, another participant from the same group, A1, was the one whose PunCAT time was the lowest of all (24.7%; Figure 2). What makes her use of PunCAT exceptional is that she hovered over no more than three or four nodes for each pun and did not click on a single node (Figure 5). She did explore a handful of source and target words she fed herself into PunCAT, but not more than most other participants. Asked to describe her experience with the tool, she said, "It was interesting. I might need to use it more often, so it gets second nature, though. It was still quite foreign to me and I was faster translating without it." During Session 2, she did most of her research on the Internet (20.9%). She also spent roughly the same amount of Internet time during Session 1 (18.4%), by far the lowest number, which indicates that she tends to rely extensively on her own internal linguistic knowledge. This assumption is also confirmed by her statement in Q1, "I came up with quite the ideas on my own." She also said that none of her target puns made use

of a term found through the tool. While the data confirm this statement for Puns 4 and 6, in the case of Pun 5 one of the three German terms she hovered over (*Welthit*: "worldwide hit") might well have prompted the term *Schlager* (a general colloquial term for a pop song and/or a particular type of German popular music), which she then entered herself into PunCAT for further exploration and eventually used in her target text. In her final version, she combined it with the German *Schläger* (meaning "bat" or "club"), which did not come up in her PunCAT search but is part of the original pun's semantic field. A gloss of her translation reads,

> "We will go look for a *Schlager*—"
>
> "I've got a *Schläger*," Lias said proudly. "Got a nail in it."
>
> "I meant a *Schlagerclub*," Glod said.
>
> "Still got a nail in it at night."

In other cases, too, the participants' personal impression of the role PunCAT played for their decision-making was not borne out by the data. A3, for instance, also said in her Q1 that she did not find the two target versions for Pun 6 through the tool, when, in fact, the words she used in both, the phonetically similar nouns *Zeit* ("time" or "tense") and *Zelt* ("tent"), were among the nodes she accessed. (She hovered over 167 nodes and clicked on 53.)

A1's reluctance to use the tool might have had to do with her confidence in her own personal resources and, as she mentioned in Q1, her unfamiliarity with the tool. Other participants also commented on their unfamiliarity with the tool. A2, for instance, said in her Q1 that she found work with the tool more stressful than without because she was not used to it. However, she found this was counterbalanced by the tool's functionality:

> [It] did provide useful input and even if I didn't choose one of the offered options/ translations, it made me think in different directions than I usually would have.

Her process data confirm her personal impression. Working on Pun 5, for example, she explored the semantic fields of the different meanings of the original "club", and we can assume that its sense of "an association of members for some common purpose" that came up in her PunCAT search, together with "club house", by extension led her to consider "membership", an avenue which, as she stated in Q1, she might not have taken without the tool's help. Her translation reads in an English gloss,

> "We will go look for a club—"
>
> "Am in a club," Lias said proudly. "Got a membership card, too."
>
> "I meant a nightclub," Glod said.
>
> "Am a member at night, too."

Group B

As mentioned above, the participants of Group B spent, on average, slightly less time in PunCAT than Group A. The participant who spent less time in PunCAT than anyone else from this group was B1 (31.8% of her Session 2 time, with 58.6% in Word and 9.6% on the Internet; Figure 2).

Regarding Pun 1, B1 made use of the same punning words as another participant who worked with the tool (B3) as well as A4, who translated this pun without PunCAT. (At least one published translation of the book also used the same words.) Their punning words, *Lehre/leer*, show

a high degree of phonetic similarity, the first term being a noun meaning "teaching" and the second an adjective meaning "empty", which is semantically close to the English source verb "lessen". Both participants who worked with the tool spent less time on the pun than A4 (17m46s; Figure 3), with B3 even producing two target versions. A gloss of B1's version, for example, would read, " 'That's why it's called *Lehre*,' the Gryphon explained, 'because the schedule gets more *leer* by the day.' "

B2, who came up with a different solution for Pun 1, just like B1, did not use any Internet resources for this pun but relied exclusively on PunCAT. She quickly zoomed in on the term *Kurs* ("course") offered by the tool, subsequently exploring some verbs in the semantic neighbourhood of *verkürzen* ("shorten"). For her final target text she then used the phonetically very close pair *Kurs/kurz* ("course"/"short"). The adjective *kurz* as such was not directly provided by PunCAT, but we can assume that the related verbs she accessed were enough of a prompt to make her come up with the adjective herself. The fact that she spent only 4m08s (Figure 3) on the translation of Pun 1, the fastest of all participants, seems to indicate that she was soon satisfied with this solution, and indeed her target pun can be seen as a valid translation, given its semantic correspondence with the original and the phonetic similarity. Interestingly, her answers in Q1 seem to somewhat contradict this conclusion, as she stated that she found translating Pun 1 "difficult" and was "not satisfied" with her target text.

After B1 had translated Pun 1 only drawing on PunCAT, she consulted various Internet resources for Puns 2 and 3. For Pun 2, she searched online for words rhyming with the German word for "enemy" (*Feind*); the greatest part of her research time, however, was spent in PunCAT (7m53s, compared to 1m09s on the Internet). One candidate pair she retrieved from PunCAT was the phonetically similar *Niere/verlieren*, the first word meaning "kidney", the second being the verb "to lose". In her final version, she retained *Niere* but replaced the verb with the nominalized form of its antonym (*Gewinnen*: "winning"), thus losing the phonetic overlap but still preserving the original semantic fields of medicine and battle: "They had a joke about *Gewinnen* and *Niere*."

B2, who had not been that satisfied with her translation of Pun 1, found translating Pun 2 "very easy" and pronounced herself "very satisfied" with the result. Translated back into English it reads, ". . . they joked that they had more intimate things behind them than an enema." Here she has made a subtle but effective pun on *hinter* ("behind"), which is used in both a figurative/temporal and a physical sense. All other participants who worked with PunCAT also managed to come up with some play on words. For example, B4 played with German prefixes in *Einlauf/Auslauf*, the first term meaning "enema", but literally "running in", and the second one having a multitude of meanings, including a run area, as in chicken-run, but also a device a fluid can "run out" of. By contrast, the only published German translation of the original story, dating back to the 1930s, unceremoniously states that "they laughed about a pun". (See Kolb [2013] for strategies used by professional translators in a previous process study.)

While several participants particularly appreciated that PunCAT pointed out the intended pun, B1, in case of Pun 3, was convinced that the tool had made a mistake, commenting in a handwritten note, "PunCAT erroneously identified picnic as the pun." Her overall experience she described as follows: "While the tool wasn't bad and I enjoyed working with it, my personal choice of pages (dictionaries, rhyming pages etc.) proved to be more useful." Her impression is not quite borne out by the actual data, which indicate that in case of Puns 1 and 2, her solutions were not prompted by any Internet resource she accessed. In the case of Pun 3, no conclusions can be drawn as she focussed all her efforts not on the intended pun but on other parts of the passage, which also explains the extremely short time of 16 seconds that she devoted to the "picnic"

phrase.

Regarding Pun 3, most participants, like B1, opted for either a direct translation, using the German word *Picknick* (the English idiom having become a truly translingual phrase and quite common in German), or worked with *Honigschlecken* or *Zuckerschlecken*, which literally mean "licking honey" or "licking sugar" but have the same double meaning as "picnic". Compared with her colleagues, B5 explored a wider range of possibilities, including the semantic fields of "play" and "meal(time)", and also terms such as "rescue" (suggested by the broader scene). She also jotted down notes and candidate pairs on paper, from which we can conclude that she was envisioning a pun involving the German homophones *Mal/Mahl* ("occasion"/"meal") and she did make use of them in one of her two target versions. The other target version used a non-punning colloquialism.

User satisfaction, translation strategies, and translation quality

Participants in both groups appreciated PunCAT's support with brainstorming and stimulating creative thinking, and felt that it reduced the level of stress. For example, B2 said in her Q1:

> For me, translating without the tool was more stressful. Even though I didn't use the exact candidates proposed by PunCAT, the tool made it a lot easier to come up with ideas. . . I used the tool mostly for inspiration. It felt like assisted brainstorming.

However, for some participants the experience was less positive, especially for B3, who explained that,

> I felt like it limited my thinking. There are so many directions you could think in but the tool only gives you synonyms. . . Working with the tool stressed me, when I had my own ideas because I felt like my mind was going to be biased.

All told, the participants produced a total of 62 target texts, of which 32 were produced with PunCAT and 30 without. Regarding translation strategies, PUN→PUN was chosen slightly more often with the tool than without it (25 vs. 21 translations). That this strategy was used so frequently probably also has to do with the study's setting and the fact that participants were encouraged to produce puns whenever possible. This necessarily entails that not all target puns can be considered successful solutions that fit all aspects of the broader context or scene, and in some cases the participants might have used different strategies in a non-academic setting (such as PUN→NON-PUN or PUN→ZERO). In both modes, the strategy PUN→NON-PUN was used in only four instances. In all other cases, participants made use of related rhetorical devices that could also achieve a whimsical effect (PUN→PUNOID), such as assonance, alliteration, or homoeoteleuton. Interestingly, two participants used irony as a rhetorical device when working without the tool on Pun 2, while this device was never used when they worked with PunCAT.

The target texts were evaluated by three external evaluators, who are experienced literary translators and teachers of literary translation. They were asked to rate the target texts using a three-level scale (fully acceptable, acceptable with some reservation/need for some revision, not acceptable). While we did expect some measure of disagreement, it still came as a surprise that only one target text (produced with PunCAT, incidentally) was rated by all as fully acceptable; seven were rejected by all as unacceptable (four produced with PunCAT, three without). Interrater reliability (Landis and Koch 1977) between two of the evaluators was fair (Cohen's $\kappa = 0.24$), while it was very poor between each of these two evaluators and the third, whose evaluations were in general much less favourable ($\kappa = 0.00, -0.28$).

Out of the 32 target texts produced with PunCAT, Evaluator 1 considered fifteen solutions fully acceptable or even successful solutions that generated a humorous effect comparable to that of the original and also fit the broader context or scene, ten as partially acceptable, with the potential of developing a satisfactory solution with some revision, and seven as not acceptable; the respective numbers for Evaluator 2 are thirteen, eleven, and eight, while Evaluator 3 fully accepted only four solutions, rating five as partially acceptable and rejecting 23.

The numbers for the mode without the tool (a total of 30 target texts) are comparable, with Evaluator 1 rating again fifteen solutions as fully acceptable, seven as partially acceptable, and eight as not acceptable. Evaluator 2 fully accepted ten solutions, accepted thirteen with reservations, and rejected seven. Evaluator 3 again rejected the highest number (nineteen), fully accepting only three solutions, with eight being rated as partially acceptable.

Conclusion

Following Low (2011, 64, 59), who argued for a "combined exploration and intuition" approach as "a systematic way to proceed instead of just waiting for inspiration", PunCAT provides users with a specialized environment intended to structure the pun translation process without unduly constraining it. Our user study appears to bear this out: We find good evidence that PunCAT can effectively support the translation process in terms of facilitating brainstorming, stimulating creative thinking, providing inspiration, and broadening the translator's pool of solution candidates by opening up larger semantic fields than traditional dictionary searches. That said, the study also shows that working styles and processes differ considerably between individuals, and PunCAT might be more suitable for some working styles than others. This may also go some way towards explaining why participants disagreed over the utility of the tool; while some valued the open exploration it afforded, others considered this feature a hindrance. A further impediment noted by the participants was their unfamiliarity with PunCAT, though this was not unexpected given the experimental setting.

The participants managed to come up with creative and valid solutions for all six puns, but in some cases ran up against gaps in the coverage for both languages' lexical-semantic networks. Regarding the future development of PunCAT, two functions stand out as particularly salient and desirable: the integration of rhyming dictionaries and/or similar resources that allow users to more easily explore and retrieve phonetically matching terms (our data showing that participants tended to focus on phonetic pairs rather than semantic pairs), and the automatic location and interpretation of punning words in the source material. This latter function, which was only simulated in PunCAT in our study, emerged as one of the features participants appreciated most.

The triangulation of software logs, questionnaires, participants' notes and target texts provided a robust basis to trace the users' interaction with PunCAT. Similar future experiments could use different setups and include, for instance, concurrent or retrospective verbalization protocols or eye-tracking software to dig even deeper into the process. Recruiting professional translators as participants may also yield somewhat different findings, though intersubject differences in working styles are also prevalent among professionals (Kolb 2019). In bringing together NLP and cognitive approaches, we also aimed to answer the clarion call that the development of computer aids for translators take more account of the users' actual working processes and practical needs (O'Brien 2020), and we consider the (further) integration of the two fields as a promising way forward to support translation in general and this rather exceptional class of translation problems in particular.

Acknowledgements

This research was funded in whole, or in part, by the Austrian Science Fund (FWF) M 2625-N31. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. The authors thank Máté Lajkó for development work on PunCAT.

Notes

1. We have released the source code for PunCAT at https://punderstanding.ofai.at/puncat under a free software licence.

References

Angelone, Erik. 2010. "Uncertainty, Uncertainty Management and Metacognitive Problem Solving in the Translation Task". In *Translation and Cognition,* edited by Gregory M. Shreve and Erik Angelone, 17–40. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/ata.xv.03ang.

Delabastita, Dirk. 1996. "Introduction to the Special Issue on Wordplay and Translation". *The Translator* 2 (2): 1–22. https://doi.org/10.1080/13556509.1996.10798970.

Fellbaum, Christiane, ed. 1998. *WordNet: An Electronic Lexical Database.* Cambridge: MIT Press.

Hamp, Birgit, and Helmut Feldweg. 1997. "GermaNet—A Lexical-Semantic Net for German". In *Proceedings of the Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications,* edited by Piek Vossen, Geert Adriaens, Nicolleta Calzolari, Antonio Sanfilippo, and Yorick Wilks, 9–15. Stroudsburg: Association for Computational Linguistics.

Hempelmann, Christian F., and Tristan Miller. 2017. "Puns: Taxonomy and Phonology". In *The Routledge Handbook of Language and Humor,* edited by Salvatore Attardo, 95–108. New York: Routledge. https://doi.org/10.4324/9781315731162-8.

Henrich, Verena, and Erhard Hinrichs. 2010. "GernEdiT—The GermaNet Editing Tool". In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010),* edited by Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, 2228–2235. Paris: European Language Resources Association.

Jiang, Jay J., and David W. Conrath. 1997. "Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy". In *Proceedings of the ROCLING X (1997) International Conference: Research in Computational Linguistics,* 19–33. Taipei: Association for Computational Linguistics and Chinese Language Processing.

Jiménez-Crespo, Miguel A. 2020. "The 'Technological Turn' in Translation Studies: Are We There Yet? A Transversal Cross-Disciplinary Approach". *Translation Spaces* 9 (2): 314–341. https://doi.org/10.1075/ts.19012.jim.

Kay, Martin. 1980. *The Proper Place of Men and Machines in Language Translation.* Working Paper CSL-80-11. Palo Alto: Xerox PARC.

Klitgård, Ida. 2018. “Wordplay and Translation”. In *The Routledge Handbook of Translation and Linguistics,* edited by Kirsten Malmkjær, 233–248. New York: Routledge. https://doi.org/10.4324/9781315692845-16.

Knospe, Sebastian, Alexander Onysko, and Maik Goth, eds. 2016. *Crossing Languages to Play with Words: Multidisciplinary Perspectives.* Part III: Translation of Wordplay. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110465600.

Kolb, Waltraud. 2013. “‘Who Are They?’: Decision-making in Literary Translation”. In *Tracks and Treks in Translation Studies: Selected Papers from the EST Congress, Leuven 2010,* edited by Catherine Way, Sonia Vandepitte, Reine Meylaerts, and Magdalena Bartlomiejczyk, 207–221. Amsterdam/Philadelphia: John Benjamins.

Kolb, Waltraud. 2019. “‘It Was On My Mind All Day’: Literary Translators Working from Home—Some Implications of Workplace Dynamics”. In *Translation Practice in the Field: Current Research on Socio-cognitive Processes,* edited by Hanna Risku, Regina Rogl, and Jelena Milosevic, 25–41. Amsterdam/Philadelphia: John Benjamins.

Kondrak, Grzegorz. 2000. “A New Algorithm for the Alignment of Phonetic Sequences”. In *1st Meeting of the North American Chapter of the Association for Computational Linguistics,* 288–295. Stroudsburg: Association for Computational Linguistics.

Landis, Richard J., and Gary G. Koch. 1977. “The Measurement of Observer Agreement for Categorical Data”. *Biometrics* 33 (1): 159–174. https://doi.org/10.2307/2529310.

Low, Peter Alan. 2011. “Translating Jokes and Puns”. *Perspectives* 19 (1): 59–70. https://doi.org/10.1080/0907676X.2010.493219.

Miller, Tristan. 2019. “The Punster’s Amanuensis: The Proper Place of Humans and Machines in the Translation of Wordplay”. In *Proceedings of the Second Workshop on Human-Informed Translation and Interpreting Technology,* 57–64. Shoumen: Incoma. https://doi.org/10.26615/issn.2683-0078.2019_007.

Miller, Tristan, Christian F. Hempelmann, and Iryna Gurevych. 2017. “SemEval-2017 Task 7: Detection and Interpretation of English Puns”. In *Proceedings of the 11th International Workshop on Semantic Evaluation,* 58–68. Stroudsburg: Association for Computational Linguistics. https://doi.org/10.18653/v1/S17-2005.

Moorkens, Joss, Antonio Toral, Sheila Castilho, and Andy Way. 2018. “Translators’ Perceptions of Literary Post-Editing Using Statistical and Neural Machine Translation”. *Translation Spaces* 7 (2): 240–262.

Nord, Christiane. 2018. *Translating as a Purposeful Activity: Functionalist Approaches Explained.* Second edition. New York: Routledge.

Novak, Josef Robert, Nobuaki Minematsu, and Keikichi Hirose. 2016. “Phonetisaurus: Exploring Grapheme-To-Phoneme Conversion with Joint *n*-Gram Models in the WFST Framework”. *Natural Language Engineering* 22 (6): 907–938. https://doi.org/10.1017/S1351324915000315.

O’Brien, Sharon. 2020. “Translation, human–computer interaction and cognition”. In *The Routledge Handbook of Translation and Cognition,* edited by Fabio Alves and Arnt Lykke Jakobsen, 376–388. London: Routledge.

Reiß, Katharina, and Hans J. Vermeer. 1984. *Grundlegung einer allgemeinen Translationstheorie.* Berlin: De Gruyter. https://doi.org/10.1515/9783111351919.

Taivalkoski-Shilov, Kristiina. 2019. "Ethical Issues Regarding Machine(-Assisted) Translation of Literary Texts". *Perspectives* 27 (5): 689–703.

Toral, Antonio, and Andy Way. 2018. "What Level of Quality Can Neural Machine Translation Attain on Literary Text?" In *Translation Quality Assessment: From Principles to Practice,* edited by Joss Moorkens, Sheila Castilho, Federico Gaspari, and Stephen Doherty, 263–287. Cham: Springer.

Vandaele, Jeroen. 2011. "Wordplay in Translation". In *Handbook of Translation Studies,* edited by Yves Gambier and Luc Doorslaer, vol. 2, 180–183. Amsterdam/Philadelphia: John Benjamins.

Youdale, Roy. 2019. *Using Computers in the Translation of Literary Style: Challenges and Opportunities.* New York: Routledge. https://doi.org/10.4324/9780429030345.