# PAUSES AS INDICATORS OF COGNITIVE EFFORT IN POST-EDITING MACHINE TRANSLATION OUTPUT[1]

**SHARON O'BRIEN**

School of Applied Language and Intercultural Studies
Dublin City University
Glasnevin, Dublin 9, Ireland
Phone: +353 1 7005832
E- mail: sharon.obrien@dcu.ie

**Abstract:** In translation process and language production research, pauses are seen as indicators of cognitive processing. Investigating the correlations between source text machine translatability and post-editing effort involves an assessment of cognitive effort. Therefore, an analysis of pauses is essential. This paper presents data from a research project which includes an analysis of pauses in post-editing, triangulated with the Choice Network Analysis method and Translog. Results suggest that the pause-to-keyboarding ratio does not differ significantly for sentences deemed to be more suitable for machine translation than for those deemed to be less suitable. Also, results confirm the finding in research elsewhere that pause duration and frequency is subject to individual differences. Finally, we suggest that while pauses provide some indication of cognitive processing, supplementary methods are required to give a fuller picture.

**Key words:** Post-editing, machine translatability, pauses, cognitive processing, Choice Network Analysis, Translog

## 1. INTRODUCTION

In both spoken and written language production research, it has been claimed that pauses are indicators of cognitive processing (Foulin 1995; Schilperoord 1996; Cenoz 2000). This claim has been adopted by translation process researchers (Séguinot 1989a, 1989b; Jakobsen 1998, 2002; Krings 2001; Hansen 2002; Alves 2006). It would seem logical then to assume that pauses are also indicators of cognitive processing in the activity known as "post-editing", i.e. the task of fixing errors in a text that has been translated from one language into another by an automatic translation system (henceforth machine translation system or MT). Research currently being undertaken by the author into the

relationship between source text machine translatability and post-editing effort, of which cognitive effort is one part, therefore includes the parameter of pauses.[2]

Machine translation (MT) involves the automatic translation of a source language text into a target language. This is a difficult task for a computer to perform, as evidenced in the literature on machine translation (cf. Trujillo 1999; Hutchins and Somers 1992; and Somers 2003 for a discussion of the main challenges). To eliminate the ambiguity inherent in natural language texts, rules can be applied. These rules are referred to as "controlled language" rules (CL). For example, a common CL rule for English as a source language is: *Avoid the use of the passive voice*. Another common rule is: *Avoid the use of the gerund*. In the domain of CL and MT, the passive voice and gerund are called "negative translatability indicators" (or NTIs), i.e. linguistic features that can have a negative impact on the MT output. By removing such indicators, it is assumed that the MT output will be better and that the effort required to post-edit a text to publishable quality will be reduced (cf. Bernth and Gdaniec 2002; Allen 2003).

The data presented here are drawn from a study on NTIs and post-editing effort that includes an analysis of temporal effort (i.e. the time required to post-edit the two sentence types) and technical effort (i.e. the level of key-boarding required to post-edit the two sentence types). In this paper, we focus on cognitive effort only. For a broader discussion of the methodologies that are suitable for measuring technical, temporal and cognitive effort in post-editing, please see (O'Brien 2005).

An investigation of pauses as potential indicators of cognitive effort in post-editing raises some interesting questions. For example, can we identify differences in pause activity between two sentence types, where one contains NTIs and the other has no NTIs? What does an analysis of pause activity for these two sentences tell us about source text translatability? Does pause analysis provide us with more information about cognitive processing in post-editing than do other methods such as Choice Network Analysis (Campbell 1999, 2000; Campbell and Hale 1999)? This paper presents exploratory research that attempts to answer these questions or, at a minimum, to identify hypotheses for further investigation.

In Section 2 of this paper, research on pauses in post-editing and translation is outlined. Section 3 discusses pause research in other domains. Section 4 presents and discusses data from the current research project and Section 5 discusses the findings.

## 2. OVERVIEW OF RESEARCH INTO PAUSES
## IN TRANSLATION AND POST-EDITING

Given that little research has been carried out to date on the activity of post-editing, it is not surprising that only one researcher (Krings 2001) comments on the use of pauses in post-editing. Krings (2001:304) states that language production research shows that pauses are of great value in the identification of processes, and especially process boundaries. In addition, the high operationality of pauses is an advantage for data analysis. He uses pauses as markers for identifying "writing acts" in post-editing activity. For Krings, the duration of a pause is one second, which, he admits, is an arbitrary unit, but he justifies this by saying that it made sense for his data analysis: one second was long enough to identify a distinguishable gap in verbalisation flow and pauses of that length were easy to identify acoustically and to record with relatively reliable intersubjectivity (Krings: 2001:210).

Some attention has also been given to the analysis of pauses in translation process studies (e.g. Séguinot 1989a, 1989b; Jakobsen 1998, 2002; Hansen 2002, Alves 2006). Séguinot, who defines pauses as "interruptions in the typing of translation" and hesitations as "unusually slow typing" (1989b:31), investigates the translation process by concentrating exclusively on the analysis of pause and hesitation phenomena. Unfortunately, she does not mention how pauses are measured. According to Séguinot, pauses typically occur

- at the end of the sentence/paragraph
- between independent clauses
- before/after subordinate clauses
- before phrases
- before subject and predicate
- at end of line/word level
- before/in words.

Jakobsen (1998) investigates pauses in the context of translation process analysis using the Translog tool.[3] He states that "the assumption that time delay during text production and translation correlates with cognitive processing is strongly supported by the systematic syntagmatic distribution of delays" (ibid:100). In his article describing how the keyboard monitoring software Translog records pauses, he claims that a pause unit of 0.20 seconds brings us close to many subjects' typing speed. He also suggests that a pause length of 1 second is appropriate for observing delays in a text production event:

> For the purpose of observing the distribution of longer delays in a text production event, a representation with a 1 second time unit will often turn out to be very appropriate because it represents all the delays we want to identify and suppresses most of the delays we are not interested in (ibid:3).

On the other end of the scale, Jakobsen argues that a time delay greater than 10 seconds will identify text initial and text final delays, delays between paragraphs, and delays appearing less systematically in front of particularly difficult text segments (ibid:84). In his 2002 article, Jakobsen investigates how time is divided between translation phases across semi-professional (senior cycle students) and professional translators. The phases he identifies are the initial orientation phase, the middle drafting phase and the final revision phase. Jakobsen observes a difference between the professional and semi-professional translators in the allocation of time between the three phases. He reports that, on average, professional translators dedicated more time to the initial phase and less time to the drafting phase than semi-professional translators.

Hansen (2002) investigates two hypotheses regarding the occurrence of pauses in the translation process. Her first hypothesis is that some translators demonstrate specific pause behaviour in translation which is independent of language direction. Her second hypothesis is that there is no correlation between the position, duration and number of pauses and the quality of the translation product. She classifies pauses into *Orientierungspausen* (orientation pauses), *Kontrollpausen* (control pauses), *Binnenpausen* (internal pauses) and *Monitoringpausen* (monitoring pauses) (ibid:33). Using Translog and an empirical approach, Hansen confirms her two hypotheses: translators do demonstrate specific pause behaviour and there is no correlation between the occurrence of pauses and the quality of the product. Hansen's finding regarding individual translator behaviour is interesting because it confirms claims regarding pauses in language production research (see Section 3) and is consistent with findings from the current study (Section 4).

Alves (2006) includes pauses in his analysis of the relationship between cognitive effort, which, following Schilperoord (1996), he terms "cognitive rhythm". Alves uses a combination of pause analysis and retrospective protocols to analyse translations of one sentence by four translators. Although he does not explicitly state it, Alves assumes that pauses are indicative of cognitive effort:

> The translator showed a pattern of pauses scattered throughout the segment, perhaps an indicator of intense monitoring of the translation process and of issues related to problem-solving and decision making. (Alves 2006:9)

Alves (ibid:6) reports that, in a previous study, the cognitive rhythm of novice translators was found to be "erratic". Echoing Hansen's (2002) finding reported above, he observed "that there was no correlation between the subjects' cognitive rhythms and the type of target text rendered by them" (Alves, 2006:6). This finding appears to be confirmed in his 2006 study (ibid:8–9) where one translator produces poor quality output without pausing for a long time while another also produces poor quality output after pausing for 89 seconds.

Pauses have, of course, been the subject of investigation in simultaneous interpreting research. In his overview of research on pauses in interpreting, Pöchhacker (2004) outlines research on the simultaneity of speaking and listening in interpreting. Pöchhacker also mentions research on the number and duration of "speech bursts" or "chunks" between pauses but he finds that the "findings from such analyses are rather varied [...] given the difference in measurement techniques, language pairs, discourse types, and skill levels" (ibid:117). However, since pause behaviour in interpreting is not the focus of this paper, we will not dwell on that topic here.

## 3. OTHER RESEARCH ON PAUSES

While it is beyond the scope of this paper to provide an extensive overview of research on pauses in the domain of language production, it is perhaps worthwhile mentioning some research in that domain as evidence from a literature search illustrates that more extensive research has been carried out in this domain than in the domains of translation or post-editing and we can use the general findings to inform our observations on pause activity in post-editing.

Studies of pauses (and hesitations) in language production can be divided into those conducted in spoken and those in written language production. As mentioned above, it is generally agreed by researchers in both domains that pauses are definite indicators of cognitive processing. Foulin (1995) observes that research into written language production has been influenced by research in the oral language production domain. In the former, however, he notes that descriptive studies are plentiful whereas the development of theory has been slow, because, Foulin claims, of methodological problems (ibid:486). A central problem is the identification – in a non-obtrusive manner – of behaviour associated with language production.

Although research on written language production draws on research from oral language production, it is noted that pauses in the latter can serve many more functions than they do in the former. For example, Cenoz (2000) mentions that pauses in oral language production can have the following functions: physiological – to allow the speaker to draw a breath; cognitive – to allow the

speaker to plan; and communicative – to allow the listener to identify demarcations in the speech stream. In written language production, pauses could be used for physiological reasons – to allow the writer to take a break from writing or keyboarding – and for cognitive reasons, but it is less likely that pauses are used for communicative reasons (unless, of course, one is researching the domain of real-time written communication over the Internet where a pause may be used to allow a discourse partner to respond to what was just written).

Regardless of whether one is more interested in oral language production or written language production, there appears to be some agreement that pauses are influenced by a number of different factors. For example, Foulin reports on other studies that have demonstrated that pauses and hesitations increase as the subject matter complexity increases. The number of pauses also seems to increase as familiarity with the subject matter decreases or as the context becomes more abstract. The same tendencies are noted in both written and oral production.

Bonin and Fayal (1996) list a number of parameters that can influence pauses, e.g. knowledge of the domain, knowledge of the text recipient, text type and emotional state of the author. These findings reflect Cenoz's claim that pause and hesitation phenomena are subject to individual variation. Van Waes and Schellens (2003) found that the mode of writing influences pause behaviour. In their analysis of pauses by users who wrote using pen and paper and those who wrote using a word processor, they discovered that, for pauses longer than 3 seconds, computer writers paused 70% more often than pen and paper writers, suggesting that the use of a word processor tends to result in a more fragmented writing process (ibid:838). They also discovered differences in the location of pauses between the two modes of writing: in the pen and paper mode, 45% of pauses were spent at paragraph boundaries while only 20% of pauses were spent at this boundary for computer writers.

Foulin points to the relative lack of interest in the revision process in studies on written language production. He attributes this to the dominant idea that pauses reflect conceptual and linguistic forward planning, but points to some research (Kaufer et al. 1986) in which systematic relations between pauses and back revision were revealed. Foulin (ibid:494) suggests that an analysis of pauses needs to take into consideration whether the pause has to do with forward planning or with revision:

> En somme, des variations de la durée des pauses, principalement celles localisées en fin d'unités structurales, pourraient également survenir en fonction de ce qui précède. Dès lors, il conviendrait, d'une part de décrire et expliquer le fonctionnement de ces pauses rétrospectives et, d'autre part, de déterminer dans quelle mesure l'activité de pause en un site donné

dépend de décisions relevant de la planification du texte subséquent ou du contrôle du texte antécédent.[4] (Foulin 1995:494)

Van Waes and Schellens compared revision activity between pen and paper writers and computer writers. In their research, large numbers of short pauses occurring in rapid succession within sentences were recorded for computer writers. This leads Van Waes and Schellens to conclude that computer writers use a writing process in which planning, formulation and revision are strongly focused on relatively small units of text.

Although it is generally agreed that pauses are indicators of cognitive processing, and should, therefore, be included as a parameter in the research question described in the introduction, it is hopefully clear from this short review of literature on the subject that pauses can be influenced by a number of parameters and that current methodologies do not allow us to specify exactly what motivates a particular pause.

# 4. ANALYSING PAUSES IN POST-EDITING

As indicated, the current research is part of a project aimed at investigating the relationship between machine translatability of source texts and the effort involved in post-editing MT output. This investigation into cognitive post-editing effort does not rely solely on pauses as indicators of effort. Pause analysis is used in triangulation, along with a method known as Choice Network Analysis. In addition, the translation process analysis tool, Translog, is used (Hansen 1999, 2002). To prepare for this analysis, an English source text was edited using Controlled Language rules. The expectation was that translation into German by an MT system of sentences with minimal occurrence of negative translatability indicators would produce a relatively high level of accuracy. Conversely, it was expected that sentences where CL rules had not been applied, and thus contained more negative translatability indicators, would prove more problematic for the MT system and, consequently, for the post-editor (for more on Controlled Language see EAMT/CLAW 2003). For ease of reference, the first sentence type (with CL rules applied) will be referred to as "$S_{min-nti}$", where "min" stands for "minimal occurrence of negative translatability indicators", and the second sentence type (no CL rules applied) will be referred to as $S_{neg}$, where "neg" refers to "occurrence of one or more negative translatability indicators".

To elicit research data, nine subjects were asked to post-edit a machine translation into German of an English text which contained a mixture of segments with negative translatability indicators (e.g. long noun phrases, gerund, passive voice) or with minimal occurrence of such indicators.[5] Each

post-editor's activity was recorded using Translog and subsequently manually analysed. The analysis of post-editing activity took into account the amount of processing time spent in pausing for each segment in the text.[6] This figure is called the "pause ratio" and is calculated by dividing the total post-editing time for the segment (as recorded in seconds by Translog) by the total time in pauses for that segment (also in seconds). For example, post-editor 1 spent a total of 23 seconds editing segment 50. Of the 23 seconds, 7 were spent in pauses. The pause ratio for post-editor 1 for segment 50 is therefore 30% (23 ÷ 7). The median pause ratio for each segment has been calculated for all nine post-editors who participated in the study. *Figure 1* shows the actual pause ratio values for all $S_{(neg)}$ sentences involved in this study (104 sentences in total) and *Figure 2* shows the pause ratio values for $S_{min-nti}$ sentences (26 in total).[7]
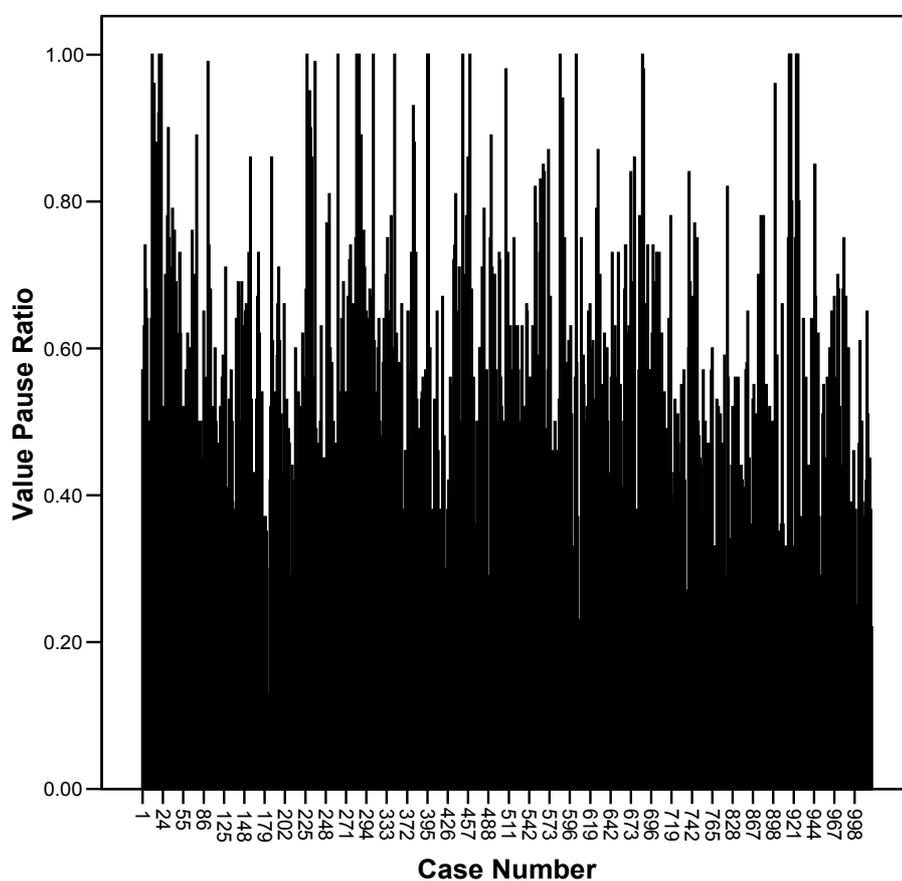


*Figure 1.* Pause Ratio Values for $S_{neg}$ sentences
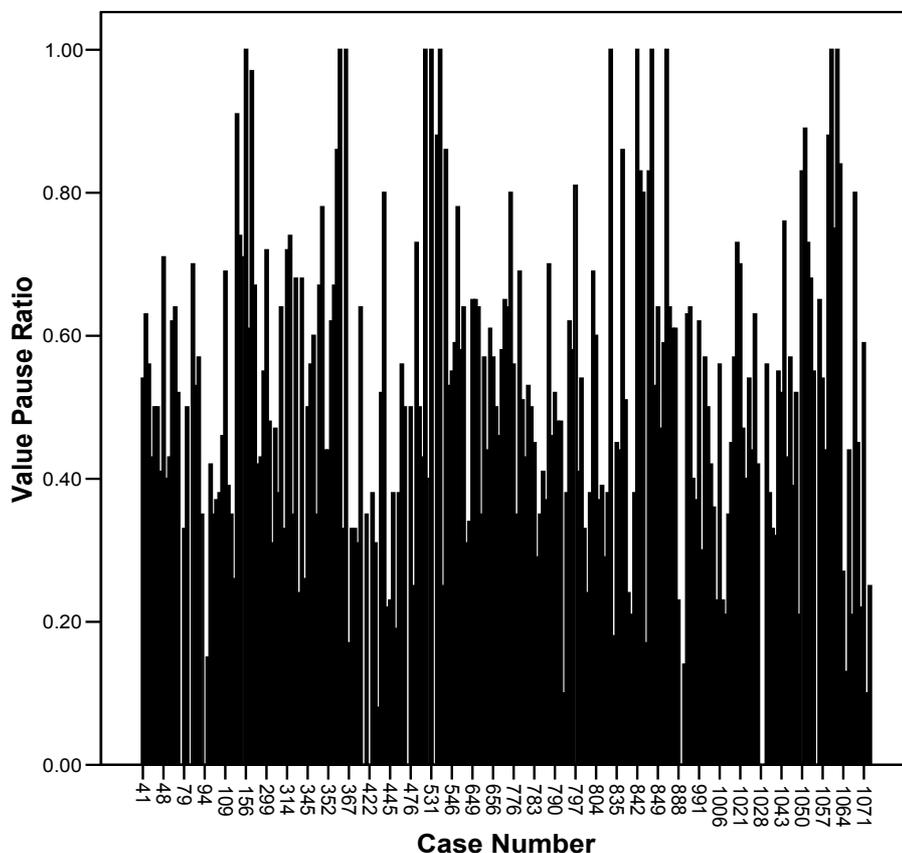
*Figure 2*. Pause Ratio Values for $S_{min-nti}$ sentences

In both cases the data suggest that the majority of values lie below the 0.50 mark (i.e. 50% of the total processing time), with some values going as high as 1.00 (i.e. 100% of the processing time). A value of 100% means that the post-editor placed the cursor on the segment to be edited, but no edits were made and the cursor was then moved to the next segment to be edited. Therefore, 100% of the processing time was spent pausing. *Table 1* shows the mean and median pause ratio values for the two sentence types.

We can see that there is only a small difference between the mean values for the two sentence types and that the median values are identical. What this means is that the median pause ratio value for both sentence types was 50%, i.e. for both sentence types post-editors paused, on average, for 50% of the total processing time. The standard deviation for $S_{(neg)}$ sentences was 0.198 and for $S_{(min)}$ sentences it was higher at 0.233.

| Sentence Type | Mean Pause Ratio | Median Pause Ratio |
|---|---|---|
| $S_{(neg)}$ | .5077 | .5000 |
| $S_{(min)}$ | .4980 | .5000 |

The data suggest that there are no significant differences for pause ratios between the two sentence types (p= 0.548). Therefore, contrary to our initial expectations, pause behaviour for post-editing sentences with minimal NTIs was not different from that of post-editing sentences containing NTIs. There are two possible explanations for this: (1) taken in isolation, pauses are not reliable indicators of cognitive load in post-editing and/or (2) there is no substantial difference in the cognitive load for post-editing for the two different sentence types. We will return to these possibilities in the Discussion in Section 5.

As already mentioned, there is general agreement that pause frequency and duration is subject to individual differences. If we take a sample of segments classified as $S_{min-nti}$ and examine the individual pause ratios across nine post-editors, we will see evidence that this is also true for post-editing (see *Figure 3*).
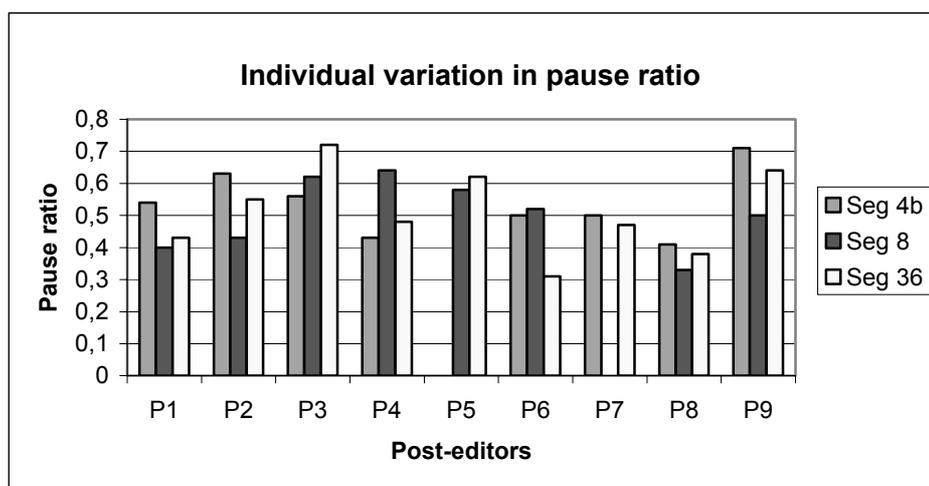


*Figure 3.* Individual Variation in Pause Ratio

*Figure 3* shows the pause ratios for post-editors P1-P9 for segments 4b, 8 and 36. All of these segments are classified as $S_{min-nti}$. These three segments were chosen for the sample because their median pause ratio values were similar in value (0.50 for segment 4b, 0.50 for segment 8, 0.48 for segment 36).

We can see from this data that both inter-subject and intra-subject variations occur, which suggests that the pause ratio differs from segment to segment for each individual subject and also from one individual subject to another across each segment.[8] This is just a small sample of the data across three segments. *Figure 4* shows the median pause ratios for the entire data set across post-editors P1–P9.
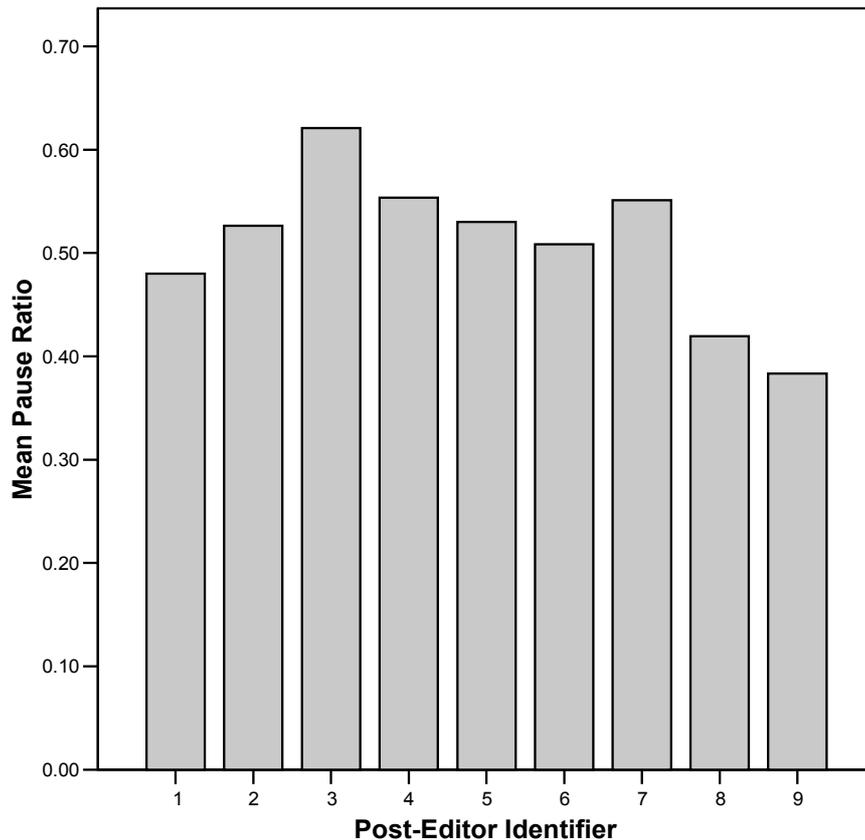


*Figure 4.* Median Pause Ratio for all Post-Editors

As we can see, the median values vary from 0.40 (or 40% of the processing time) for Subject 9, to over 0.60 (or 60% of the processing time) for Subject 3. One interpretation of this data could be that pausing is a highly individual characteristic in post-editing as well as general language production.[9]

To establish the usefulness of pause analysis as an indicator of cognitive processing in post-editing, the frequency and duration of pauses for all nine post-editors was recorded. This analysis was triangulated with results from

Choice Network Analysis. Choice Network Analysis examines the product of the post-editing process. It assumes that places where the product differs substantially over the nine post-editors represent points where the post-editors encountered most difficulty during the post-editing process. The triangulation process is highlighted here using one source sentence, which is classified as an $S_{neg}$. The sentence contains four negative translatability indicators. They are the gerund (two occurrences), a proper noun, an abbreviation and "missing" punctuation.[10] The source sentence reads: *Introduction to Using ArborText Epic for Editing SGML Text Files* The raw machine translation output in German is: *Einführung dabei, ArborText Epic für das Editieren von SGML Textdateien zu verwenden.*

The post-edited segments for each of the post-editors are represented here in *Table 2* (*a* and *b*).[11] The text marked with bold highlights those parts of the post-edited sentences that show the greatest degree of variation and, in keeping with the theory of CNA, which are thus assumed to represent the parts that the post-editors found most difficult.

As can be seen from *Tables 2a* and *2b*, the translation difficulty indicator that causes most processing here is the first gerund, *using*. On the other hand, the second gerund, *editing*, causes little or no post-editing. The first gerund is interpreted by the MT system incorrectly as a present progressive form, while the second is interpreted correctly as a gerund. Neither the proper noun *Arbour Text Epic* nor the abbreviation *SGML* have resulted in poor MT output or post-editing activity. Also, the fact that this sentence is a heading, with no full-stop, has not caused problems for the MT system or for the post-editors.

*Table 2a*

| MT | Einführung | dabei, | – | – | ArborText Epic |
|----|------------|--------|---|---|----------------|
| P1 | Einführung | in die | Verwendung | von | ArborText Epic |
| P2 | Einführung | zur | Verwendung | von | ArborText Epic |
| P3 | Einführung | in die | Benutzung | von | ArborText Epic |
| P4 | Einführung | in die | Verwendung | von | ArborText Epic |
| P5 | Einführung | in das | Editieren | – | – |
| P6 | Einführung | zur | Verwendung | von | ArborText Epic |
| P7 | Einführung | in die | Verwendung | von | ArborText Epic |
| P8 | Einführung | in die | Verwendung | von | ArborText Epic |
| P9 | Einführung | in das das | Editieren | – | – |

*Table 2b*

| MT | für das | Editieren | von SGML Textdateien | zu verwenden. | – |
|---|---|---|---|---|---|
| P1 | für das | Editieren | von SGML-Textdateien | – | – |
| P2 | für das | Editieren | von SGML-Textdateien | – | – |
| P3 | für das | Editieren | von SGML-Textdateien | – | — |
| P4 | für das | Editieren | von SGML-Textdateien | – | – |
| P5 | – | – | von SGML-Textdateien | – | mit ArborText Epic |
| P6 | für das | Editieren | von SGML-Textdateien | – | – |
| P7 | für das | Editieren | von SGML–Textdateien | – | – |
| P8 | zum | Editieren | von SGML-Textdateien | – | – |
| P9 | – | – | von SGML-Textdateien | – | mit ArborText Epic |

*Table 3*

Translog Symbols

| | |
|---|---|
| ✳ | Pause of 1 second or less |
| [✳:10.82] | Extended pause with length of time |
| [Ctrl⇦] | Jump one word to left |
| [Ctrl→] | Jump one word to right |
| [🖰:12] | Mouse movement |
| [ShftCtrl→] | Select a word to the left |
| [ShftCtrl←] | Select a word to the right |
| ⊠ | Delete to left |
| ↓ | Move down one line of text |
| ← | Move one character to the left |
| → | Move one character to the right |
| ↑ | Move up one line of text |
| in◆die◆Verwendung | Typing of these exact characters |
| ◆ | Insert blank space |
| [Dictionary:·specify·=·Not·found] | Unsuccessful dictionary look-up for the word "specify" |
| [Dictionary:·document·type·definition··=·DTD;] | Successful dictionary look-up for the term "document type definition"; equivalent given "DTD" |
| [ST↓] | Scroll source text down |
| ⇦ | Backspace |

An examination of the pause data for this segment allows us to investigate whether the evidence from Choice Network Analysis regarding post-editing effort is accurately reflected in the pause data. The pause data are extracted from the log file produced by Translog. It is possible to change the minimum

pause measurement in Translog but the research to date on pauses in translation and post-editing uses 1 second as a minimum measurement and although, as Krings admits, 1 second is an arbitrary measurement, it was decided to continue using 1 second as a minimum measurement for this research to ensure comparability of results. A key to the symbols produced by Translog in the log file is provided in *Table 3*. The log data have been broken into "chunks" here for ease of reference using the pause marker (either ∗ or [[∗:] ) as the start of each new chunk. *Table 3* provides a key to the Translog symbols.

We will present the pause data for P1 and P4 here. The pause data for all other post-editors can be found in *Appendix 1*.

```
**[Ctrl⇦]
[*:13.62][Ctrl➙]in◆die◆Verwendung◆
**von◆⊠⊠⊠⊠⊠⊠
**⊠[Ctrl➙][Ctrl➙]
*[Ctrl➙]
**[Ctrl➙][Ctrl➙][Ctrl➙][Ctrl➙]⊠-
[Ctrl➙]⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠➙
⇓
```

*Figure 5.* Keyboard activity for Post-Editor 1

```
*[Ctrl⇦]
[*:55.22][⇡]
**⇦
***in◆die◆Verwednung⇦⇦⇦⇦ung◆von◆
*⊠⊠⊠⊠⊠⊠⊠⊠⊠➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙
*⇦⇦⇦⇦⇦⇦⇦⇦⇦⇦⇦⇦⇦⇦⇦⇦⇦⇦⇦⇦⇦n➙➙➙➙➙
➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙
***➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙
*⇦⇦⊠➙➙➙➙➙➙➙➙➙➙⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠
*********
```

*Figure 6.* Keyboard Activity for Post-Editor 4

P1 pauses six times during the post-editing of this segment. Four of these pauses last for two seconds, one for one second and one for an extended period of 13.62 seconds. Of the six pauses, two are followed by cursor activity, i.e. jumping one word to the right or left, two are followed by the addition of new text (i.e. typing of *in der Verwendung* and of *von* followed by the deletion of six letters), and two are followed by a combination of cursor movement and

deletions. The explicit introduction of *in die Verwendung* corresponds to the part of the text identified as being difficult in the Choice Network Analysis. It is interesting that this is also preceded by the longest pause. However, we should contrast this with the data for P4.

P4 pauses nine times altogether. The pauses vary in duration from one second to 55.22 seconds. While the typing of *in die Verwednung* (sic) is also represented here, it is not preceded by the pause of longest duration (which precedes a mouse click). We can see from the data that the activity following pauses again varies from text addition, to text deletion, to cursor manipulation.

The pause data presented here are generated by two post-editors for one segment only. However, a visual analysis of the pause data over all nine post-editors suggest that the two sets of data presented here are good representations of pause behaviour in the sample – they contain both short and long pauses (the latter being greater than 10 seconds); they record both character insertion and deletion and they record other keyboarding activity, such as the use of the arrow and function keys. The findings suggest the following:

- Those aspects of the source text that are identified as being "difficult" in Choice Network Analysis are always preceded by a pause.
- However, there appears to be little correspondence between the duration of the pause and editing of the "difficult" element. In other words, the difficult element might be preceded by a pause of one second, while chunks of cursor activity might be preceded by pauses of much longer duration. This suggests that either solutions for problematic translations can be found in very short pause times or that the cognitive effort required to find those solutions is not necessarily expended in the pauses directly preceding the implementation of those solutions.
- We can see from the data provided by Translog for P4, that the post-editor uses the cursor to scan the words of the MT output, without making any edits (see multiple occurrences of "→"). It is possible that P4 is using this technique in lieu of a pause, i.e. as the cursor runs across each word, the word or phrase is being considered by the post-editor.

## 5. DISCUSSION

This paper describes some exploratory research into the occurrence of pauses during post-editing of machine translation output. The data presented here have raised some interesting hypotheses on pauses in post-editing and support some of the findings on pauses in language production and translation research.

Our data suggest that correlations between pause duration and location and the occurrence or extent of text edits in post-editing are difficult to establish.

Here we recall similar findings by Hansen (2002) that there is no correlation between the occurrence of pauses and product quality and Alves (2006) that cognitive rhythm and "contextual effect" can sometimes have a positive relationship, but can sometimes also have a negative relationship.

We have demonstrated that the pause ratios for $S_{min-nti}$s and $S_{neg}$s do not differ substantially. This suggests that the same proportion of cursor or keyboarding activity is taking place for sentences classified as $S_{min-nti}$ as for sentences classified as $S_{neg}$. As we stated previously, there are two possible explanations for this: (1) taken in isolation, pauses are not reliable indicators of cognitive load in post-editing and/or (2) there is no substantial difference in the cognitive load for post-editing for the two different sentence types. In the case of the first explanation, we know that it is generally accepted that pauses are indicators of cognitive effort. The problem lies with linking the duration and location of pauses to cognitive processing.[12] In the case of the second explanation, data on technical and temporal effort will provide us with further insight into the effort required for the two sentence types. It is our intention to correlate the data on cognitive effort with technical and temporal effort in a future publication. To have the same level of textual edits for both sentence types would be surprising as this would have negative implications for the use of Controlled Language rules which are used to improve machine translation output and, consequently, to reduce post-editing effort.

It appears that some post-editors use the arrow keys to navigate the text word for word. It is suggested here that this activity could be a substitute for pausing and that it is akin to moving a pen or a finger under printed words when reading or revising. When this activity is ongoing, it is likely that cognitive processing is also ongoing.

Choice Network Analysis identifies those parts of a sentence that are most changed during post-editing. It is assumed that for these changes to be implemented, cognitive effort is required. However, it appears that Choice Network Analysis tells us only part of the story. The analysis of pause data reveals that pauses do not just occur in "difficult" parts of the sentence and that pauses are not always followed by text edits. Pause analysis, therefore, supplements Choice Network Analysis.

Finally, we have re-confirmed the finding in language production and translation research that the patterns of pause activity vary from one individual to the next.

## 6. CONCLUSIONS

Clearly the potential exists for more research on this topic. A general conclusion is that pauses are somewhat useful as indicators of cognitive effort in post-editing, but, as with language production research, it is very difficult to correlate

cognitive effort with pauses, source text difficulty and target text quality. This is in keeping with findings to date in translation process research. For the question regarding the correlation between source text machine translatability and post-editing effort, pauses help in the measurement of this effort, but other methods such as Choice Network Analysis and keyboard monitoring using Translog are also important for supplementing pause data. A question that remains to be answered is this: is it that there are no correlations between pause duration and position or is it that we need more studies on pause behaviour translation and post-editing in order to discover correlations? We do not wish to dismiss pause analysis as a potential indicator of translation difficulty or cognitive rhythm. We would like to see more studies in this domain in order to clarify how useful pause analysis really is for translation, or post-editing, research.

## Notes

[1] The author wishes to acknowledge the Irish Research Council for the Humanities and Social Sciences (IRCHSS) and IBM Research for their support.

[2] Explaining the notion of machine translatability in detail is beyond the scope of this paper. The term refers to a measurement which indicates the expected success of machine translation for a particular source text. For further information on this topic, see Bernth and Gdaniec 2002.

[3] A tool developed by Jakobsen in the Copenhagen Business School which records keyboard activity during translation.

[4] In summary, variations in the duration of pauses, especially those located at the end of a structural unit, could also arise from what went before. Therefore, one ought to describe and explain the function of these pauses on the basis of what preceded them and, in addition, determine to what extent pause activity in a particular location is a function of forward planning or text revision. [my translation]

[5] All post-editors were qualifed translators with several years of translation experience with this text type and in this domain (user manuals and IT). Subjects were asked to fill out a questionnaire on their qualifications and experience prior to selection. The data were then used to select subjects who were as homogenous as possible in terms of qualifications and experience.

[6] A "segment" can be a sentence or a heading, a table entry or an entry in a bulleted list. The data presented here are drawn from a text containing 165 segments in total (1 777 words).

[7] The reason for the difference in numbers of sentence types is that the focus of this research was on trying to establish what correlations, if any, exist between post-editing effort and specific NTIs and it was necessary to have each NTI represented a number of times. The reader is reminded that each sentence was processed by nine different post-editors. The $S_{min-nti}$ sentences acted as control sentences. "Case number" in each figure refers to a unique identifier given to each sentence in the Statistics Package, SPSS.

[8] The missing data points for P5 and P7 result from the fact that these post-editors ignored segments 4b and 8 respectively.

[9] Nevertheless we acknowledge that the data could be interpreted in a different way, i.e. the post-editors demonstrate homogenous behaviour by pausing on average between 40 and 60% of the processing time. There is an element of subjectivity here in the interpretation of the data. The author is treating the difference of 20% as evidence of individuality in pause behaviour.

[10] In fact, the punctuation (full-stop) is not missing since this segment acts as a heading. Nonetheless, this feature could be problematic for some MT engines so it is included here.

[11] Table 2 is broken into two parts (2a and 2b) for ease of reference. Table 2b displays the latter half of the sentence.

[12] We are aware that Think-Aloud Protocol (TAP) has been used by researchers to gain insight into cognitive processing during translation. However, the utterances produced during either simultaneous or retrospective protocols are not always reliable. See (O'Brien 2005) for more discussion of this.

# References

Allen, J. 2003. Post-editing. In: Somers, H. (ed.) *Computers and Translation – A Translator's Guide.* Amsterdam: John Benjamins, 297−318.

Alves, F. 2006. A Relevance-Theoretic Approach to Effort and Effect in Translation: Discussing the Cognitive Interface between Inferential Processing, Problem-Solving and Decision-Making, In: *Proceedings of the International Symposium on New Horizons in Theoretical Translation Studies*, 19-20 January, Department of Translation, Hong Kong: The Chinese University, 1−12.

Bernth, A. & Gdaniec, C. 2002. "MTranslatability". *Machine Translation* Vol 16. No. 3. 175−218.

Bonin, P. & Fayol, M. 1996. L'Etude en temps réel de la production du langage écrit, pourquoi et comment, *Etudes de Linguistique Appliquée*, Vol 101. Jan–Mar, 8−19.

Campbell, S. 1999. A Cognitive Approach to Source Text Difficulty in Translation. *Target* Vol.11. No.1, 33−63.

Campbell, S. 2000. Choice Network Analysis in Translation Research. In: Olohan, M. (ed.), *Intercultural Faultlines: Research Models in Translation Studies: Textual and Cognitive Aspects.* Manchester: St. Jerome. 29−42.

Campbell, S. & Hale, S. 1999. What Makes a Text Difficult to Translate? *Proceedings of the 1998 ALAA Congress*, Available at: http://www.cltr.uq.edu.au/alaa/proceed/camphale.html, last checked on: February 14, 2006.

Cenoz, J. 2000. Pauses and hesitation phenomena in second language production. *ITL-Review of Applied Linguistics* Vol. 127−128, 53−69.

EAMT/CLAW 2003. Proceedings of the Joint Conference Combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop, Dublin: Dublin City University.

Foulin, J-N. 1995. Pauses et débits: Les indicateurs temporels de la production écrite, *L'Année Psychologique* Vol. 95. 483−504.

Hansen, G. 1999. *Probing the process in translation: methods and results.* Copenhagen Studies in Language. Copenhagen: Samfundslitteratur.

Hansen, G. 2002. Zeit und Qualität im Übersetzungsprozess, Hansen, G. (ed.) *Empirical Translation Studies: Process and Product.* Copenhagen Studies in Language 27. Copenhagen: Samfundslitteratur. 29−54.

Hutchins, J. & Somers, H. 1992. *An Introduction to Machine Translation.* London: Academic Press.

Jakobsen, A-L. 1998. Logging Time Delay in Translation, LSP Texts and the Translation Process. *Copenhagen Working Papers.* 73−101.

Jakobsen, A-L. 2002. Translation Drafting by Professional Translators and by Translation Students, In: Hansen, G. (ed.) *Empirical Translation Studies: Process and Product,* Copenhagen Studies in Language 27. Copenhagen: Samfundslitteratur. 191−204.

Kaufer, D. S., Hayes, J. R. & Flower, L. S. 1998. Composing Written Sentences. *Research in the Teaching of English* Vol. 20. 121–140.

Krings, H. P. 2001. Repairing Texts: Empirical Investigations of Machine Translation Post-editing processes. In: Koby, G. S. (ed.) Kent, Ohio: Kent State University Press.

O'Brien, Sh. 2005. Methodologies for Measuring the Correlations between Post-editing Effort and Machine Translatability. *Machine Translation* Vol. 19. No.1. 37–58.

Pöchhacker, H. 2004. *Introducing Interpreting Studies*, London: Routledge.

Schilperoord, J. 1996. It's About Time: Temporal Aspects of Cognitive Processes in Text Production. Amsterdam: Rodopi.

Séguinot, C. (ed.) 1989a. *The Translation Process.* School of Translation, York University. Toronto: HG Publications.

Séguinot, C. 1989b. The Translation Process: An Experimental Study. In: *The Translation Process.* School of Translation, York University. Toronto: H.G. Publications. 21–53.

Setton, R. 1999. *Simultaneous Interpretation: A Cognitive-Pragmatic Analysis.* Amsterdam and Philadelphia: John Benjamins.

Somers, H. (ed.) 2003. *Computers and Translation: A Translator's Guide.* Amsterdam: John Benjamins.

Trujillo, A. 1999. Translation Engines: Techniques for Machine Translation. London: Springer-Verlag.

Van Waes, L. & Schellens, P.J. 2003. Writing Profiles: the Effect of the Writing Mode on Pausing and Revision Patterns of Experienced Writers. *Journal of Pragmatics* Vol. 35. 829–853.

# APPENDIX

## Translog Data for All Nine Post-editors For Segment 1

**P1**

\*\*[Ctrl⇦]
[\*:13.62][Ctrl➙]in◆die◆Verwendung◆
\*\*von◆⊠⊠⊠⊠⊠⊠
\*\*⊠[Ctrl➙][Ctrl➙]
\*[Ctrl➙]
\*\*[Ctrl➙][Ctrl➙][Ctrl➙][Ctrl➙]⊗[Ctrl➙]⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠➙↓

**P2**

[Ctrl⇦]\*\*\*\*\*\*\*\*[ST↑][ST↑]
\*[↶🖰]
\*⇧⇧
\*\*↑[Ctrl➙][Ctrl➙][Ctrl⬅]
\*\*zur◆Verwendung◆
\*von◆
\*\*[ShftCtrl➙]⊠[Ctrl➙][Ctrl➙]
\*[Ctrl➙][Ctrl➙][Ctrl➙][Ctrl➙][Ctrl➙]⊗
\*\*[Ctrl➙][ShftCtrl➙][ShftCtrl➙]⊠\*

**P3**

\*[Ctrl⇦]
\*\*\*\*\*[Ctrl⇦]
\*\*\*\*\*[Ctrl➙]
\*\*\*\*\*in◆die◆Benutzung◆vn◆[Ctrl⊠]⊗⊗on⊠[Ctrl➙][Ctrl➙]\*\*\*\*[Ctrl➙][Ctrl➙][Ctrl➙]
[Ctrl➙][Ctrl➙]
\*[Ctrl➙][Ctrl⊠][Ctrl⊠]
\*[Ctrl⬅]⬅⊠

**P4**

\*[Ctrl⇦]
[\*:55.22][↶🖰]
\*\*⬅
\*\*\*in◆die◆Verwednung⬅⬅⬅⬅ung◆von◆
\*⊠⊠⊠⊠⊠⊠⊠⊠➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙
\*⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅n➙➙➙➙➙➙➙➙➙➙➙
➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙
\*\*\*➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙➙
\*⬅⬅⊠➙➙➙➙➙➙➙➙➙➙➙⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠
\*\*\*\*\*\*\*\*\*

**P5**

\*[Ctrl⇦]
[\*:15.29][Ctrl➙]

\*\*\*in٠das٠Edti
\*⌫⌫itieren٠von٠
\*\*\*[Ctrl→][Ctrl←]SGMLTextdateien٠mit٠
\*[ShftCtrl→]⌦٠⌫[Ctrl→][Ctrl→]
\*\*\*[Ctrl⌦][Ctrl⌦][Ctrl⌦][Ctrl⌦][Ctrl⌦][Ctrl⌦]
\*\*\*[Ctrl⌦][Ctrl⌦]→→

**P6**

\*[Ctrl⇦]
[\*:19.31][Ctrl→][ShftCtrl→][Shft←]
\*\*\*\*\*\*zur٠Verwendung٠von
\*→→→→→→→→→→→→→→→
\*\*\*\*\*→→→→→→→→→→→→→→→→→→→→→→→
\*\*\*\*⌫\*→→→→→→→→→→→→→→→←←←[ShftCtrl→][ShftCtrl→]⌦
\*\*\*

**P7**

\*[Ctrl⇦]
[\*:15.84][Ctrl→]
\*\*\*in٠die٠VErwe⌫⌫⌫⌫erwendung٠von٠⌦⌦⌦⌦⌦⌦⌦
\*\*\*\*[Ctrl→][Ctrl→][Ctrl→][Ctrl→][Ctrl→][Ctrl→][Ctrl→][Ctrl→]⌦⌦⌦⌦⌦⌦⌦⌦⌦
⌦⌦

**P8**

\*[Ctrl⇦]
\*\*\*\*\*\*\*→→→→→→→→→→→→
\*←
\*i
\*n٠
\*\*\*die٠Verwendung٠von
\*[Shft→][Shft→][Shft→][Shft→][Shft→][Shft→]
\*⌦→→→→→→→→→→→→→→→→→→zum٠[Shft→][Shft→][Shft→][Shft→][Shft→][Shft
→][Shft→]⌦⌦
\*→→→→→→→→→→→→→→→→→→-
⌦⇨[Shft←][Shft←][Shft←][Shft←][Shft←][Shft←][Shft←][Shft←][Shft←][Shft←][Shft←]
[Shft←][Shft←]⌦

**P9**

\*[Ctrl⇦]
\*\*\*\*\*\*\*\*→→→→→→→→→→
\*\*in٠d
\*ie⌫
\*\*\*⌫as٠
\*\*[Ctrl⌦]⌦[ShftCtrl→][ShftCtrl→][Cut]
\*⌦⌦⌦⌦[Ctrl→][Ctrl→][Ctrl→][Ctrl→][Ctrl→]←←[Ctrl←]←⌦[Ctrl→]
\*[Paste:··ArborText·Epic·][Ctrl←][Ctrl←]mit٠[Ctrl→][Ctrl→]←[ShftCtrl→][ShftCtrl→]⌦