



# Cognitive Explorations of Think-Aloud Protocols: Insights from Eye Tracking

Stephen Doherty  
*SALIS Research Showcase*  
November 9<sup>th</sup> 2011

# Overview

- Think-Aloud Protocols
- TAPs in Research
- As a Research Method
- Focus on Two Studies
- Conclusions and Recommendations

# Think-Aloud Protocols

- Description:
  - Data elicitation technique in which a participant is asked to verbalise his/her own thoughts during or after performing a task
  - Adopted from psychology; developed primarily by Ericsson and Simon (1980, 1987, 1993)
  - Often synonymous with the introspective method
- Types of TAPs:
  - Concurrent/retrospective
  - Moderation on a scale of active to inactive

# Think-Aloud Protocols

- When content is unavailable or inaccurate, participants still think-aloud; “unaware of their unawareness” (Wilson *et al.* 2008)
- Less valid in everyday or unusual circumstances (White 1988), e.g. research environment?
- The ‘introspection illusion’ explains cognitive biases (Pronin 2007).
- Participants see their own introspections as reliable, whereas other people are judged on their overt behaviour (Pronin & Matthew 2007)
- People generally see themselves as less biased than others (Pronin & Matthew 2007)
- Content from thinking-aloud can be confabulated (Nisbett & Wilson 1977)

# Think-Aloud Protocols

- Thinking-aloud during decisions making is especially affected by bias (Johansson *et al.* 2006)
- Questioning, e.g. probes, hugely impact the validity of concurrent TAPs, but less so for retrospective Taylor and Dionne (2000)
- TAPs adopted in translation process research (e.g. Krings 1986, 1987, 1988)
- Jääskeläinen (1999): TAPs unable to illustrate objects of translators' attention; they reflect only a limited portion of cognitive processing
- Jakobsen (2003): concurrent TAPs significantly increased the time it takes to produce a translation; professionals had far fewer verbalisations, i.e. automated processing

# As a Research Method

- On the Fence:
  - Individual differences
  - Impact of the researcher, e.g. white-coat effect, additional cognitive load (Preece 1994)
  - Researcher-participant relationship
  - Cultural influences
  - Research aims – is validity important?

# As a Research Method

- Strengths:
  - Valid measure of human subjective human experience
  - Resource-cheap
  - Portability
  - Rich qualitative data
  - Participants can reflect on findings in retrospective method, e.g. education (Bartels 2008)
  - Provides data that cannot be captured by other means, e.g. human-computer interaction, usability

# As a Research Method

- Weaknesses:
  - Human experience cannot be reduced to TAPs (Nielsen *et al.* 2002)
  - Participant behaviour does not adequately reflect higher cognitive processing (Whiteside *et al.* 1993)
  - Forgetting, fabrication (Russo *et al.* 1989, Fleck & Weisberg 2004)
  - Limitation of short-term memory (Eysenck & Keane 2008)
  - Rejection in some fields, e.g. cognitive psychology
  - Implicit knowledge (Broadbent *et al.* 1986)
  - Non-experts and cognitive effort, the TAP burden (Gile 1998)
  - Automated processing in professionals (Jakobsen 2008)
  - Unconscious cognitive processing (Eger *et al.* 2007)
  - Cognitive processing much faster than verbal processing (Eger *et al.* 2007)
  - Situation-dependent - the everyday and the research lab
  - Priming studies

# Focus on Two Studies

- Aim of Studies:
  - Pilot - to test the methodology and validated eye tracking as a valid method of translation evaluation (machine translation - MT)
  - Main - to investigate the impact of a controlled language on the MT output in terms of readability and comprehension
  
- Methodology:
  - Eye Tracking
  - Retrospective TAP (Pilot)
  - Post-task evaluation (Main study)
  - Recall test (Main study)
  - Other textual metrics
  
- Materials & Procedures
  - 12 then 24 native speakers of French
  - Reading/evaluating translations translated from English to French
  - Technical support documentation

# Focus on Two Studies

- Results from Study A:
  - Evaluators showed a sufficiently moderate level of inter-rater agreement ( $\kappa = .63$ ); agreed with original evaluators
  - Eye tracking metrics mixed but sig. correlations found
  - No sig. correlation between verbalisations and translation quality
  - However, sig. correlation between 'Silence' and 'good' quality ( $r = .52, p = .05$ )
  - Coding?
- Results from Study B:
  - Human evaluation more sensitive to controlled language changes
  - Automated metrics less so
  - Eye-tracking metrics mixed but sig. correlations found
  - Compensation for errors?

# Outlook & Recommendations

- No perfect method in research involving humans (?)
- Identification of limitations in the context of study
- Use as a supplement to other methods
- Qualitative/quantitative research
- For example, mixed-methods triangulation...
- Reporting of results

# References/Readings:

Dumas, J.S. (2001) "Usability Testing Methods: Think-Aloud Protocols," in Design by People For People: Essays on Usability, UPA, pp 119-129

Dumas, J.S. & Redish, J.C., (1993) A Practical Guide to Usability Testing. Norwood, NJ, Ablex Publishing Corp.

Ericsson, K.A. & Simon, H.A. (1984, 1993) Protocol analysis: Verbal reports as data (Rev. ed). Cambridge, MA: MIT Press

Eysenck, M. W. and Keane, M. T. 2010. Cognitive Psychology: A Student's Handbook (6th ed.). East Sussex and New York: Psychology Press.

Gilhooly, K. J. Individual differences in thinking-aloud performance. *Current Psychology*, 5, (4), 328-334.

Hannu, K., & Pallab, P. (2000). "A comparison of concurrent and retrospective verbal protocol analysis". *American Journal of Psychology* (University of Illinois Press) 113 (3): 387-404.

Jääskeläinen, R. 1999. Tapping the Process: An Explorative Study of the Cognitive and Affective Factors Involved in Translating. PhD thesis. University of Joensuu.

Jakobsen, A. L. 2003. Effects of think aloud on translation speed, revision and segmentation. IN: Alves, F. (ed.). *Triangulating Translation. Perspectives in Process Oriented Research*. Amsterdam: John Benjamins, pp. 69-95.

Johansson, Petter; Lars Hall, Sverker Sikström, Betty Tärning, Andreas Lind (2006). "How something can be said about telling more than we can know: On choice blindness and introspection". *Consciousness and Cognition* (Elsevier) 15 (4): 673-692

# References/Readings:

Kuniavsky, M. (2003) *Observing the User Experience: A Practitioner's Guide to User Research*, San Francisco: Morgan Kaufmann Publishers, Inc.

Nielsen, J., (1993) *Usability Engineering*. Chestnut Hill, MA: Academic Press, Inc.

Nielsen, J., Clemmensen, T., & Yssing, C., (2002) "Getting access to what goes on in people's heads? - Reflections on the think-aloud technique", paper presented to NordiCHI, Aarhus, Denmark, October 19-23

Nisbett, Richard E.; Timothy D. Wilson (1977). "Telling more than we can know: Verbal reports on mental processes". *Psychological Review* 8: 231-259., reprinted in David Lewis Hamilton, ed (2005). *Social cognition: key readings*. Psychology Press.)

Preece, J. (1994), *Human-Computer Interaction*, Addison-Wesley, England

Pronin, Emily; Matthew B. Kugler (July 2007). "Valuing thoughts, ignoring behaviour: The introspection illusion as a source of the bias blind spot". *Journal of Experimental Social Psychology* (Elsevier) 43 (4): 565-578.

Tamler, H. (2001) "How (Much) to Intervene in a Usability Testing Session," in *Design by People For People: Essays on Usability*, UPA, pp 165-171

White, Peter A. (1988). "Knowing more about what we can tell: 'Introspective access' and causal report accuracy 10 years later". *British Journal of Psychology* (British Psychological Society) 79 (1): 13-45.)

Whiteside, J., Bennett, J.L., & Holtzblatt, K., (1988) "Usability Engineering: Our Experience and Evolution," in *Handbook of Human Computer Interaction*; edited by Helander, M. New York, NY: Elsevier Science Publishers

Wilson, Timothy D.; Yoav Bar-Anan (August 22, 2008). "The Unseen Mind". *Science* (American Association for the Advancement of Science) 321 (5892): 1046-1047.

# End

- Questions?
- Thanks!
- Your experiences...