Linguistic complexity and cognitive workload: measurement and management

Asad B. Sayeed and Vera Demberg
Multimodal Computing and Interaction Cluster of Excellence
Saarland University
66123 Saarbrücken, Saarland, Germany
{asayeed,vera}@coli.uni-sb.de
http://www.coli.uni-saarland.de/~{asayeed,vera}

Course type Introductory

Abstract This proposal describes a course in issues related to the effect of language and linguistic complexity on human cognitive workload. We propose this with an eye towards both practical issues for the implementation of spoken dialogue systems and and the greater scientific understanding of the interaction of language and the other components of cognition. With rapid expansion of spoken interfaces on smartphones and cell phones, this is a new issue of increasing urgency, particularly in areas such as driver distraction, However, it has an existing body of background work in multitasking experimentation and linguistic complexity measurement. We will present experimental and computational techniques in measuring, analysing, and modeling linguistic workload effects. We will finally discuss these issues in terms of human-computer interaction and user interface design.

1 Introduction

We are proposing a course that presents the ideas and concepts behind the measurement and modeling of cognitive workload accruing from linguistic interaction, with a particular gaze towards spoken dialogue systems. A combination of technological advances makes this topic not only increasingly relevant in terms of applications, but increasingly interesting in terms of increasing our scientific understanding of human language processing. Our course will be a broad introduction to the history, current state of research, and research techniques in this area.

Including such famous examples as Apple's Siri and Google Glass, interaction with automated systems through the use of human language continues to grow. Task attention therefore comes to matter to the usability of spoken interfaces, and thus cognitive workload is now a relevant yet little-taught topic that will be increasingly relevant to linguistics and informatics students who are interested in future work in a variety of systems and applications.

However, the proliferation of these systems also presents an opportunity for expanding our understanding of how language works in the context of the mind as a whole. How does the mind manage multiple sources of input and multiple modes of output? There is considerable work in psychology on multitasking for other modes of interaction. However, workload measurement techniques have reached the point where it is now time to assess how the specific characteristics of language fit in to working memory and attentional cognition.

We intend for this course to plant the seeds in the minds of attendees to consider the human-computer interaction and multitasking aspects of language technology. We will do this within the context of language modeling, workload measurement, and psycholinguistic experimentation.

2 Language technology in the human environment

The major focus of language and workload research focuses on the task of driving, where major safety issues come into play in an area of pervasive social importance. The literature contains results showing that using a cell phone (hand-held or free speaking) during driving can impair driving performance, and also some more in-depth analysis suggesting that concentration on the interaction with the remote conversational partner can lead to "inattention blindness" (Strayer et al., 2003), i.e. that cell phone conversation reduces the attention to visual inputs and therefore reduces driving performance.

There are many instances of this kind of research. Recarte and Nunes (2003) found that passive listening had little effect on driver distractedness, but performing a mental task that requires a response ("production task"), had a significant effect on drivers' visual search and decision-making capabilities. *Contra* Recarte and Nunes, Kubose et al. (2006) find that their comprehension and production tasks have equally deleterious effects on driving ability.

Beyond cars, this line of research has implications for other areas of human-computer linguistic interaction. Adapting the level of linguistic complexity is not only relevant for multi-tasking during driving, but also for human-computer interaction in critical situations, for example (Kruijff, 2012), in rescue robots used in catastrophic scenarios such as car accidents in a tunnel, earth quake rescue, etc. and 2) should exhibit a similar communication behavior, and some of the same measures of linguistic complexity might be relevant in these scenarios.

We will cover this genre of work, but also cover much more recent efforts in language and workload measurement, such as Demberg et al. (2013). This more recent work finds a very direct relationship between the moment-to-moment processing of linguistic input and physiological measures of cognitive workload.

3 Cognitive workload

What is cognitive workload for the purpose of this course? It is a way of characterizing the "mental effort" required to complete a task or set of tasks. There is today a general consensus that mental resources are limited, and that these resources need to be distributed over concurrent tasks. Distribution of limited resources over concurrent tasks can lead to decreased performance on some or all of these tasks as a consequence. These concepts of resource distribution and gradual effort go back to the late 60s (Moray, 1967; Kahneman, 1973) and different versions of the idea are now implemented in several state-of-the-art processing models (Baddeley, 2003; Wickens, 2008; Just et al., 2003). Given the resource limitation, we can narrow down the definition of cognitive load to refer to the relative proportion of mental resources that are consumed by the task set in question. Participants in this course will be given an overview of these theoretical and experimental background works.

4 Workload measurement

How is cognitive workload measured experimentally? We will explain some of the major techniques of measurement. One increasingly prominent measure is pupil area. Laeng et al. (2012) have shown convincing evidence that pupil area is closely related to the activation of the noradrenergic system of the brain, which is centered around the locus coeruleus (LC) on the brainstem. The LC is closely connected to the brain areas known to be involved in selective attention processing, and neuronal activation studies have shown that pupil dilation is closely time-locked to the firing rates of LC neurons.

Pupil size has long been shown to reflect a linguistic burden (Engelhardt et al., 2010), including for relatively fine-grained linguistic phenomena (Just and Carpenter, 1993). A growing variant of pupil area is the patented Index of Cognitive Activity (ICA) (Marshall, 2002), which is derived from pupil area in order to capture the small fluctuations of the pupil. Engonopoulos et al. (2013) and Demberg et al. (2013) show that these small fluctuations can be used to detect workload from linguistic complexity. Engonopoulos et al. (2013) in particular demonstrate the other side of workload measurement: task performance, in this case in terms of driving deviation from a pre-set course (Mahr et al., 2012).

An older method of workload measurement is galvanic skin conductance, which is an electrophysiological measures known to correlate well with cognitive load and is used in the driving context (Mehler et al., 2009). Skin conductance increases due to the activation of sweat glands under conditions of stress or cognitive load; it is possible to distinguish emotional stress from mental effort (Setz et al., 2010).

We will present not only the history and techniques to the course participants, but also analysis techniques such as linear mixed-effects regression modeling (Bates and Sarkar, 2007) and other statistical data exploration methods.

5 Linguistic complexity

We will present material on the above topics to the course in the context of efforts to "drill deeper" into the fine structure of linguistic communication. One stream of research into the processing characteristics of language, known as Dependency Locality Theory (DLT), occupies itself with the algorithmic costs of storing and integrating items into mental representations (Gibson, 2000). Another stream of research, known as the Uniform Information Density (UID) Hypothesis, focuses on the principles of management of information (phonological, syntactic, semantic, discursive) during language-based interaction (Levy, 2008; Demberg et al., 2012). During the course, we will introduce these approaches to explaining well-known psycholinguistic effects (e.g. Bader and Meng, 1999) and relate these to the above approaches to workload, in the end raising the future research questions about how to approach these issues in practical systems (Demberg and Sayeed, 2011).

References

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews: Neuroscience*, 4:829–839.

Bader, M. and Meng, M. (1999). Subject-object ambiguities in german embedded clauses: An across-the-board comparison. *Journal of Psycholinguistic Research*, 28(2).

Bates, D. and Sarkar, D. (2007). *lme4: Linear mixed-effects models using S4 classes*.

Demberg, V., Kiagia, E., and Sayeed, A. (2013). Language and cognitive load in a dual task environment. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society (CogSci-13)*.

Demberg, V. and Sayeed, A. (2011). Linguistic cognitive load: implications for automotive uis. In *Cognitive load and in-vehicle human-machine interaction, workshop at AutomotiveUI 2011*.

Demberg, V., Sayeed, A. B., Gorinski, P. J., and Engonopoulos, N. (2012). Syntactic surprisal has an effect on spoken word duration. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP 2012)*, Jeju, Korea.

Engelhardt, P. E., Ferreira, F., and Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *Quarterly journal of experimental psychology*, 63:639—645.

Engonopoulos, N., Sayeed, A., and Demberg, V. (2013). Language and cognitive load in a dual task environment. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society (CogSci-13)*.

Gibson, E. (2000). Dependency locality theory: A distance-dased theory of linguistic complexity. In Marantz, A., Miyashita, Y., and O'Neil, W., editors, *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*, pages 95–126. MIT Press, Cambridge, MA.

Just, M. and Carpenter, P. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 47(2):310.

Just, M. A., Carpenter, P. A., and Miyake, A. (2003). Neuroindices of cognitive workload: neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical issues in ergonomic science*, 4(1-2):56–88.

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall Inc.

Kruijff, G. (2012). How could we model cohesiveness in team social fabric in human-robot teams performing under stress? In *2012 AAAI Spring Symposium Series*.

Kubose, T. T., Bock, K., Dell, G. S., Garney, S. M., Kramer, A. F., and Mayhugh, J. (2006). The effects of speech production and speech comprehension on simulated driving performance. *Applied cognitive psychology*, 20(1):43–63.

Laeng, B., Sirois, S., and Gredebäck, G. (2012). Pupillometry: a window to the preconscious? *Perspectives on psychological science*, 7(1):18–27.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.

Mahr, A., Feld, M., Moniri, M., and Math, R. (2012). The ConTRe (Continuous Tracking and Reaction) task: A flexible approach for assessing driver cognitive workload with high sensitivity. In *Automotive User Interfaces and Interactive Vehicular Applications*, pages 88–91.

Marshall, S. (2002). The index of cognitive activity: Measuring cognitive workload. In *Human Factors and Power Plants, 2002. Proceedings of the 7th Conference on*, pages 7–5. IEEE.

Mehler, B., Reimer, B., Coughlin, J. F., and Dusek, J. A. (2009). Incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation research record*, 2138:6—12.

Moray, N. (1967). Where is capacity limited? a survey and a model. *Acta psychologica*, 27:84–92.

Recarte, M. A. and Nunes, L. M. (2003). Mental workload while driving: effects on visual search, discrimination, and decision-making. *Journal of experimental psychology: applied*, 9(2):119–137.

Setz, C., Arnrich, B., Schumm, J., LaMarca, R., Troster, G., and Ehlert, U. (2010). Discriminating stress from cognitive load using a wearable eda device. *IEEE transactions on information technology in biomedicine*, 14(2):410–417.

Strayer, D., Drews, F., and Johnston, W. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of experimental psychology: Applied*, 9(1):23.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors*, 50(3):449–455.

Tentative outline

- Day 1: Wider context and foundations.
  - Spoken dialogue systems in practice—current status, recent advances.
  - Potential problems from multitasking (e.g. driving with a cell phone, user error, etc.).
  - Neurocognitive workload basics—models of attentional cognition.
- Day 2: Linguistic complexity
  - What can make native language input difficult?
  - Quantiying the burden of language: information-theoretic and algorithmic approaches.
  - Relevant rudiments of language modeling techniques, esp. syntactic and semantic.
  - Tools and resources.
- Day 3: Workload measurement and experimental technique
  - Physiological measures: skin conductance, pupillometry, fMRI, etc.
  - Task-oriented measures: e.g. driving deviation, language interaction task performance.
  - Traditional tasks (e.g. digit memory) and results.
  - Data analysis (regression, time-series auto-correlation).
- Day 4: Language and workload
  - Coarse-grained experimentation on language in multitasking environments.
  - The fine structure of grammar vs. fluctuations in measured workload.
- Day 5: Issues for language interfaces
  - Natural language generation.
  - Information presentation and dialogue management.
  - Bringing it all together: user-adaptive dialogue systems.

Prerequisites

This is an introductory course, so advanced knowledge is not expected. A basic understanding of statistics is expected. Understanding of grammar, parsing, and/or spoken dialogue systems will help, but is not required.

Funding

Organizers are presently located within Germany (Saarbrücken), so only local travel support (rail) and accommodation is required.