# Eye-tracking during concurrent scene and sentence presentation

Eric Auer <eric@coli.uni-sb.de>

Diploma Thesis at Saarland University,
Faculty 4 – Humanities II,
Department 4.7 – General Linguistics.

**Thesis Committee:**
Pia Knoeferle
Matthew Crocker

6th September 2005

# Contents

# List of Figures

# List of Tables

**Abstract**

Eye-tracking has a long tradition in psycholinguistic research, for example helping to investigate human language comprehension in the presence of non-linguistic context. We present a concurrent scene and written sentence presentation study, where we manipulate congruence and sentence type of structurally ambiguous German sentences. The corresponding displayed scenes show role-ambiguous characters involved in depicted actions. We investigate the effects of depicted events on ambiguity resolution in the context of a congruence decision task.

Our concurrent presentation experiment combines aspects of several recent studies on scene-sentence integration. We put our results in context and present a comprehensive analysis of our data. The detailed analysis allows direct comparison with various aspects of existing and future results of other studies. We replicate some results from related studies and discuss parallels and differences in the results. We present some interesting interactions between sentence type and congruence effects.

Another goal of our study is to evaluate the new open source eye-tracking environment NEWTRACK (running on FreeDOS) and our custom portable DPIFilter post-processing and visualization software. NEWTRACK is more suitable for our study than for example PCEXPT. We give an overview of the features of NEWTRACK (e.g. support for near-standard hardware and automatic stimulus image analysis) and assess the current performance. We also review the current limitations of NEWTRACK and make suggestions for future enhancements.

## Zusammenfassung

Eye-tracking hat eine lange Tradition in der psycholinguistischen Forschung, zum Beispiel in Untersuchungen zum menschlichen Sprachverstehen im Zusammenhang mit nichtlinguistischem Kontext. Wir präsentieren eine Studie mit gleichzeitiger Anzeige von Text und Bild, in der wir Kongruenz und Satzbau lokal strukturell ambiger deutscher Stimuli manipulieren. Die verwendeten Bilder zeigen Personen und Aktionen, wobei ambige Rollenzuordnungen vorkommen. Wir untersuchen die Effekte der im Bild gezeigten Ereignisse auf die Ambiguitätsverarbeitung im Kontext der Aufgabe, über die Übereinstimmung von Bild und Text zu entscheiden.

Unser Experiment verbindet Eigenschaften mehrerer aktueller Studien zur Text / Bild Integration. Wir präsentieren unsere Ergebnisse im Kontext und analysieren unsere Daten umfassend. Die detaillierte Analyse erlaubt den direkten Vergleich unserer Ergebnisse mit den Ergebnissen bestehender und zukünftiger Experimente. Wir reproduzieren einige Ergebnisse ähnlicher Studien und besprechen Parallelen und Unterschiede in den Ergebnissen. Wir präsentieren einige interessante Wechselwirkungen zwischen Satzbau und Kongruenzeffekten.

Ein weiteres Ziel unserer Studie ist der Test der neuen open source eye-tracking Umgebung NEWTRACK (unter FreeDOS) und der plattformunabhängigen DPIFilter Nachbearbeitungs- und Visualisierungs-Software dieses Lehrstuhles. NEWTRACK eignet sich aus verschiedenen Gründen besser für unsere Studie als zum Beispiel PCEXPT. Wir bieten einen Überblick über die Funktionen von NEWTRACK (zum Beispiel Unterstützung aktueller Standard-Hardware und automatische Analyse der Stimulus-Bilder) und geben eine Beurteilung der aktuellen Leistungsfähigkeit. Wir betrachten auch die Grenzen der aktuellen NEWTRACK Version und schlagen Verbesserungen für zukünftige Versionen vor.

## Acknowledgements

The author wishes to thank the thesis committee for their helpful and personal backing throughout the writing of this thesis: Matt Crocker, for his support in the development and the adoption of NEWTRACK, a new modern eye-tracking software. It has been fun to work for the psycholinguistics department before, but the new software is a particularily contenting project. Thanks to Pia Knoeferle, for her help in the design of the experiment, for making available proven stimulus components from earlier experiment and for always being available for comprehensive feedback on this thesis. Some other members of the department have been a big help as well, they know who they are. Even before it had a psycholinguistics department, the computational linguistics faculty at Saarland University has always been an inspiring place to learn and work at. Thanks also to Ralf Brown for writing his comprehensive and freely available "Interrupt List", which is a big help with PC programming. Thanks to Shravan Vasishth, and others for their inspiring and helpful comments and to Ruth Kusterer for proofreading. This thesis would have been much harder to create without powerful free software like LaTeX, gnuplot, ImageMagick, Perl, Java, GCC / DJGPP, NASM, Linux and FreeDOS.

Less scientific thanks go to friends and family, for their support and because friends and real life are, after all, the most important things to have. Best wishes to the friends and fellow students for their lives at and after university. In a more virtual way, the FreeDOS community must be thanked for their challenging project, sharpening both programming and English language skills. Without all the support and stimulation from his parents to get into all kinds of technical things, the author would not be the technically skilled and educated person that he is now. Thanks a lot, and thanks for the financial support. Last but certainly not least there is a beloved best friend to be thanked for being there, being a reason to get things done and enjoy some leisure time with her then – and of course for her language lessons. It has been a pleasure to learn her language and it opened up valuable new perspectives.

# Chapter 1

# Introduction

## 1.1 Eye-tracking studies

### 1.1.1 Eye-tracking in context

Eye-trackers can be used to investigate human language processing because what people look at (the gaze position) is closely related to how they process language: People look longer at parts of displayed text if those texts differ from what they expected (garden-path effects, unusual word order, e.g. [Hem93] found related reading time effects). Another example is that hearing an utterance while looking at a related scene guides the scene inspection. Various studies suggest that this happens in a very direct, incremental way, down to the word level.

Tanenhaus et al report in [TSKES95] that participants listening to spoken instructions look at objects immediately after or even while hearing relevant words. Structurally ambiguous instructions like "put the apple on the towel in the box" caused different eye movement patterns depending on visual context (whether there was only one apple or rather two, one of them on a towel). An interesting finding of the Tanenhaus et al study is that the visual context directly affects the scene inspection strategy. So effects are not *limited* to eye movements being guided by linguistic input. Rather, effects can also happen in the other direction. Another study about the close interaction between *semantic* ambiguity resolution and visual context is [STCC99]: Hearing "tall" already made people look at a tall glass (before hearing "glass") if the scene also contained a contrasting small glass.

Not only items but also actions are affected: hearing "eats" causes *anticipatory* eye movements to the edible object in a scene, if only one such object is visible – see [AK99]. So agent, verb and world knowledge alone already cause expectations about the patient of an action, even though the patient is specified explicitly only moments later anyway. Such expectations are still practically useful as they reduce the risk of misunderstandings (e.g. in noisy environments). The study of [KAH03] presents a visual scene and spoken text: For example, a man, a girl, a motorbike, a carousel, beer and sweets are shown. The text mentions the man or the girl and either ride or taste as actions before finally mentioning the target object of the action. There are effects of the mentioned agent and effects of the mentioned verb (action). The scenes only show agents and objects, but do not actually depict the actions, so world knowledge is selecting plausible targets here. As expected, linguistic input is found to have a direct and closely time-locked effect on eye movements.

Ambiguity in language has been a focus of psycholinguistic research over decades: A short overview can be found in [Alt98]. Researchers have tried to explain ambiguity processing, parsing preferences and garden-path effects for example with hard and heuristical rules, parallel constraints, plausibility (based on world knowledge). They created models with competing interpretations (competition slowing down decision when several interpretations have similar strength / probability), rule-based models and connectionist models, for example. To compare models with human sentence processing, one wants to know which properties of their "input" humans use (and how) to resolve ambiguities and what causes difficulties for them.

Because language processing and eye movements are often closely related, eye-tracking can give interesting insights about the mental processes of human language processing to psycholinguists. An overview of the close connection between eye movements and reading and information processing and of research done on that topic is given in [Ray98].

Many eye-tracking studies can exploit the fact that human vision focus does not generally move around smoothly: Instead, the gaze position changes suddenly, in a so-called saccadic eye movement (saccade) several times per second, between steady "fixation" phases (see e.g. "How do I look?" / Bristol University re:search 2002 at www.bris.ac.uk). During fixations, a region of interest is kept at the center of the vision field. The image is at the foveal region of the retina. The foveal region is a small area with optimum visual resolution, covering only about 2 degrees of visual angle. Even for moving objects, smooth pursuit eye movements keep the object of interest locked at the center of the vision field for a while. For experiments which do not involve moving objects, the properties of human eye movements allow us to summarize text / image inspection as a short list of fixation coordinates and timespans for each trial. The gathered data can then be passed through a number of filters, removing unnecessary detail. Finally, descriptive and statistical analysis can be performed on the filtered data.

### 1.1.2 Related studies

Starting in the 1970s, psycholinguists have been studying concurrent scene and text presentation, for example to find out about effects of scene / text match and mismatch. Later studies that involved depicted scenes have often used spoken utterances along with visual scenes or have used a serial presentation paradigm. Notable exceptions are [CYG92], analyzing eye movements while understanding cartoons, and [Heg92], analyzing eye movements while understanding mechanical diagrams and their captions. Later studies about concurrent text and scene display include [RRS+01] about print advertisements and [UJR04] about road scenes shown along with text about those scenes. The latter study also compares concurrent to serial presentation of text and scene.

This thesis combines concurrent presentation of written language and images with research about effects of structural disambiguation and a comparison task: As in [UJR04], our participants had to decide whether some text and a depicted scene match. And as in [TSKES95] and [KCSP05], we investigate the interactions of language and visual context for structural disambiguation. However, the combination of factors is new: [KCSP05] did not search for effects of match / mismatch, [TSKES95] focused on referential contrast as supporting factor in syntactic processing, not on match / mismatch. While they compared ambiguous to unambiguous utterances, [TSKES95] did not investigate sentence type effects. Rather, they analyze the effect of different kinds of referential context. Moreover, the latter three studies all use spoken rather than written language.

The study conducted for this thesis combines aspects of one of the [UJR04] experiments (match / mismatch decision task in concurrent scene and sentence display) with recent visual world research: [KCSP05] use visual context with depicted events to ease structural disambiguation of initially ambiguous spoken utterances. Our study can help to generalize across modalities and provides a more detailed analysis than [UJR04].

The serial presentation experiments of [UJR04] show that image first stimuli are generally harder to process than text first or concurrent presentation stimuli. [RRS+01] also found that participants prefer to read the text before inspecting the image in depth for concurrent presentation stimuli. An interesting problem with [UJR04] is that the authors fail to show an effect of congruence in the serial presentation experiment while they found an effect in the concurrent presentation experiment. In the [KC05] study, the authors argue that the problem of [UJR04] could be due to not looking at regions of text separately. Rather, [UJR04] just analyzes overall text inspection and image inspection properties.

The [KC05] study uses unambiguous images. They compare matching, mismatching and "mismatching because no action depicted" congruence levels. In addition, the study checks for effects of sentence type (SVO, subject verb object, and OVS, object verb subject). So the study shares several aspects of our experiment – but not the concurrent presentation and not the use of ([KCSP05] style) ambiguous images. Our study preserves the local ambiguity at least until the verb is read, with sentences starting with an ambiguous noun phrase referring to a depicted character with ambiguous role in the depicted actions. The [KC05] study uses serial (image first) presentation. In that aspect, the authors create a setting similar to one of the experiments of [UJR04]. There is no explicit congruence verification task in [KC05]. One of the main [KC05] results is that congruence effects *can* be found even with serial presentation stimuli if one analyzes the per-region fixation data rather than only the overall sentence / image fixation data. As a second result, the SVO / OVS sentence type factor is found to have a strong effect on fixations on all analyzed regions.

Our study aims to replicate and extend results of [KC05]. The design of our stimuli will enable us to get insights from both sentence inspection and scene inspection data. As [KC04] argue for a close interaction of scene and linguistic context as input to our highly optimized cognitive system, a concurrent presentation experiment seems to be very appropriate. Concurrent presentation of spoken utterances and visual scenes would further improve the parallelism of the input but would make it impossible to collect data about sentence inspection.

### 1.1.3 The project

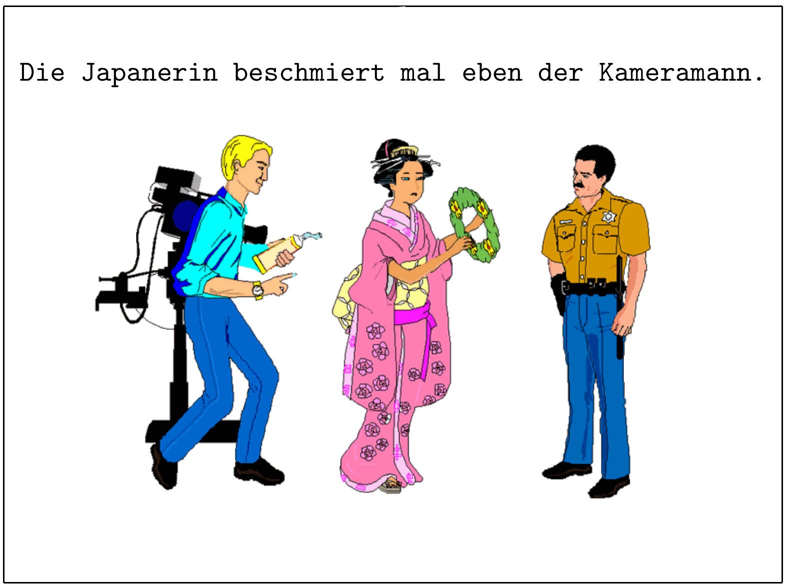
The experiment conducted for this thesis analyzes the interpretation of German SVO (subject verb object) or OVS (object verb subject) sentences and structural disambiguation of SVO

/ OVS sentences when shown along with a matching or nonmatching depicted event scene. German language allows both word orders, but OVS is dispreferred (see [Hem93]). Sentences were designed to be initially ambiguous: For female noun phrases, no ambiguity-resolving case-marking is present in the determiner. The case-marking in the final second noun phrase disambiguates the sentence later.

Each sentence is shown concurrently with an image which either does or does not depict the action described in the sentence (match / mismatch). An example stimulus is shown in figure 1.1. The character referred to by the first noun phrase is in the middle, involved in two depicted actions. So the depicted scene is ambiguous as well, but the verb of the simultaneously presented text selects one of the two depicted actions, presumably allowing the anticipation of the other character involved in the action without having to know the second noun phrase of the text. In [KCSP05], depicted (ambiguous but matching) scenes faciliated the comprehension of initially ambiguous spoken OVS utterances. Our stimuli are largely based on the [KCSP05] materials. Thanks to the authors for making them available for our experiment.

The combination of trial properties "word order" and "match / mismatch" leads to four different ways to assign roles to the three (agent, middle / ambiguous, patient) characters shown in the image for a trial. For the two factors word order (levels: SVO / OVS) and congruence (levels: match / mismatch), we expect: More fixations / longer inspection times for mismatch, and OVS comprehension being facilitated by visual context in the match case. We also expect an interaction effect of sentence type on the agent and patient character regions.

Figure 1.1: Sample stimulus: Sentence and scene are shown concurrently. The sentence is shown above the scene: "The Japanese[nom/acc] besmears (just so) the cameraman[nom]" – The cameraman (just so) besmears the Japanese.



## 1.2 The software

One of the achievements of our study is to evaluate the new NEWTRACK eye-tracking environment (see [Aue05]) and the (Java) DPIFilter post-processing (log-viewing and log-cleaning) software. We run NEWTRACK in the free DOS compatible single-tasking FreeDOS [@FREEDOS] operating system to avoid time skews caused by other tasks consuming CPU time.

NEWTRACK and DPIFilter allow us to conduct eye-tracking experiments with stimuli that contain images and text. Image-only and (multiline) text-only trials are possible with this software as well, but that has not been used for the experiment described in this thesis. One feature that was used (and saved quite a bit of work) is the ability of NEWTRACK to automatically chunk images into their depicted items as long as they are not too close to each other or contain too much background color.

The features of NEWTRACK make it more suitable for our experiment compared to PC-EXPT (see [@PCEXPT]). With the present PCEXPT software, concurrent text and image presentation is not supported. We would have had to prepare image files with the text *drawn* on the image canvas in advance, for every used combination of images and text. PCEXPT does not give feedback about fixation locations during the experiment (NEWTRACK does), and it requires several steps of postprocessing to get the region / fixation assignments. NEWTRACK has a separately useable module to automatically analyze and chunk images, to find the locations of all depicted items. NEWTRACK also stores the word locations for the stimulus text in the logs. Combining both features gives our DPIFilter software all needed information about the locations of all analyzed regions on the stimulus display.

Another advantage of NEWTRACK is that it is specifically designed to work on modern hardware and to be easy to maintain: The software is written in GNU C and the sources are split into functional units of reasonable size. The used compilers are freely available, as is the FreeDOS operating system which supports the big amounts of memory and disk space which can be found on a standard PC today. There is no need to recompile the program for each experiment: Stimulus display is controlled by a carefully designed scripting language which supports text stimuli, image stimuli and mixed stimuli.

All Pentium or newer IBM PC compatible processors are supported and no separate chronometer hardware is needed. Display output works with two normal VGA compatible PCI or newer graphics cards, but classic “VGA plus monochrome” setup is supported as well. The latter is part of a compatibility driver set built into NEWTRACK to allow use on existing PCEXPT-specific hardware, but using faster systems with two VGA graphics cards is strongly recommended. Coordinate digitizing can be done with existing PCEXPT compatible A/D boards or with a custom low-complexity device connected to the printer port. The design of the software allows easy addition of drivers for other hardware: For testing, NEWTRACK can run on single-screen computers and a computer mouse can be used to simulate the eye-tracker input. The keyboard can be used to simulate a buttonbox.

## 1.3 Overview

In this chapter, our eye-tracking study was placed in the context of previous research. We outlined that a new software suite will be used and what we expect from the software and which results we expect from the experiment.

Chapter 2 about software follows: The new real-time eye-tracking environment NEWTRACK (section 2.1) and the DPIFilter software used to clean and review the log files (section 2.2) are presented in detail.

Chapter 3 consists of several sections about the eye-tracking study conducted for this thesis. We first outline the method[1], and then provide a report of our findings (data usability, descriptive and inferential analysis), followed by a discussion of our results.

The following chapter 4 contains the general discussion part, with a review of the software performance and a summary of our experimental results in context of previous research.

The final chapter presents some conclusive remarks about software and the results of our eye-tracking study. Appendices with further details about materials and results and the list of references follow.

[1] includes information about the eye-tracking hardware and the reached accuracy

# Chapter 2

# The first study with a new system

## 2.1 NEWTRACK – a new eye-tracking software

### 2.1.1 The NEWTRACK software

NEWTRACK is a new software for eye-tracking with standard computer equipment, initially developed as a term project by Eric Auer [Aue05] in 2004 and later enhanced to support concurrent image and text presentation and other features used in our experiment. During the experiment, several improvement ideas emerged, and new versions with those improvements are now used for other experiments.

Our eye-tracking software is more suitable for our experiment than for example [@PCEXPT]: The NEWTRACK software supports concurrent text and image presentation and can work on and use the abilities of modern PC hardware. The stimulus definition script language is flexible and the log files provide detailed information in a convenient text format. The software was developed with user-friendliness of the system and good maintainability of the (open) source code in mind. It can be compiled with modern freely available compilers and can run on freely available operating systems.

NEWTRACK does not contain any eye-tracking algorithms itself. It must be used with an eye-tracker which provides at least raw (uncalibrated) coordinates directly. Those are read with help of a low-complexity custom-made adapter box which is connected to the printer port of the PC. Using a Fourward DPI eye-tracker (as described in section 3.1.3.1) has the advantage of providing the raw coordinates as voltages directly. All processing which is directly related to determining the participant's eye location is done by hardware there.

It would be possible to use camera-based eye-trackers indirectly: Most systems offer add-ons where the calibrated coordinates are made available as voltages by some add-on card plugged to the PC which contains the video processing hardware. That would, of course, require using two PCs, one for NEWTRACK and one for the camera eye-tracker. It also adds the complication that the PC with the video processing hardware would likely have to communicate with the PC which is running NEWTRACK about display contents in calibration mode. An alternative would be to immobilize the participant's head and feed only *uncalibrated* pupil location information to the NEWTRACK PC, bypassing the head location and rotation information processing on the camera eye-tracker PC[1]. The use of NEWTRACK is not necessarily limited to systems with DPI eye-trackers. However, NEWTRACK cannot do the camera data processing itself (yet). Camera-based eye-trackers often use proprietary high-speed image processing hardware for which no driver sources are available. So for now, we recommend to use a Fourward DPI eye-tracker with NEWTRACK.

Using a Fourward DPI eye-tracker to take care of the actual low-level aspects of eye-tracking in hardware, the NEWTRACK software can focus on more high-level aspects of eye-tracking experiments: It loads data about the stimulus to be displayed, establishes the actual stimulus region locations, renders the stimulus and collects and analyzes real-time data about gaze positions during the experiment. NEWTRACK also does the sequencing of events, reacts to participant (YES / NO) or operator (NEXT / CALIBRATE) buttonpresses and has a built-in calibration function, as described later in this document.

### 2.1.2 Hardware requirements, data acquisition and button box

To keep the PC hardware equipment at reasonable cost and easy to duplicate, a single standard PC (AMD Athlon 700 MHz CPU, 64 MB RAM) is used: In addition to the main AGP (Accelerated Graphics Port, a common graphics card connector / slot type), the PC has a second standard PCI (Peripheral Component Interconnect, the current standard for all sorts of extension cards) graphics card. One graphics card is used to display stimulus text and images for the participant, while the other graphics card is run in text mode to display real-time information about experiment progress for the operator, on a screen which is not visible to the participant during the experiment. Ralf Brown's Interrupt List [@RBIL] has been a great source of information about hardware and drivers during the development of NEWTRACK and the built-in dual-head display drivers.

One issue in setting up a NEWTRACK environment in your lab will be the selection of two suitable graphics cards: Either you use two identical graphics cards (which can be set up by the same bootstrapping code) or you have to take extra steps to make the second graphics cards initialize properly. As DOS does have no built-in graphics system and the BIOS only initializes one graphics card at boot time, the initialization of the second card must be done manually. If both cards are the same, the RBdualVGA approach can be used [@RBIL], using the same bootstrapping code for both cards. Our lab uses another method, where a small helper application (written in Assembly language for the NASM assembler [@NASM]) loads the bootstrapping code temporarily and removes it from memory again after initialization. This method does, however, not work for all graphics cards. For testing, when there is no second graphics card, NEWTRACK uses interlaced display output showing both participant and operator display contents on a single screen is used.

The interface between eye-tracker and PC is a small A/D (analog/digital) converter box attached to the printer port, containing two standard monolithic A/D circuits (LTC1286)

[1] Still a more powerful calibration scheme might be needed to let NEWTRACK calculate screen coordinates from raw pupil location data. Current scheme is quadratic curve best fit calibrated to a $3 \times 3$ grid.

which are operated in parallel, one for the vertical and one for the horizontal gaze position. The eye-tracker outputs the gaze position as a pair of voltages, which are quantified to raw coordinates in 0 to 4095 range, at medium speed: The LTC1286 can send up to 12.500 digitized samples per second over a serial line, but our software reads only up to 2.500 samples per second, still well above the frequency response of the eye-tracker. Technical information about the construction of the A/D converter box is available in the NEWTRACK documentation [Aue05]. NEWTRACK is open source software [@GPL], compiled with the DOS port of GNU C, DJGPP by DJ Delorie [@DJGPP]), so a good alternative to building your own LTC1286-based converter box is to port existing drivers for other A/D converters to NEWTRACK.

Apart from capturing eye movements, the software has access to the state of two buttons located on a box held by the participant. Those are also connected to the printer port here, but the joystick port would have worked as well. Using mouse- or keyboard-buttons would have the disadvantage of having a limited time resolution, because those devices communicate through low-speed serial links. The participant can give YES / NO responses by pressing the buttons. For convenience, the operator has two further buttons, labelled “next” and “calibrate”, with obvious purposes: They are used to advance to the next trial (or, during calibration, to the next gaze position) and to enter calibration mode. The latter is only possible between trials.

### 2.1.3 High-speed processing

Having to process and analyze the data and update the operator text screen reduces the achievable sample rate. Fonts are rendered by hardware in text mode, making text screen updates a generally low overhead task. Using a 700 MHz processor is enough to keep the time between two samples well below one millisecond in all cases. Sampling is stopped while the display for the participant is updated, which is a more time-consuming task. Enabling the software to show animated stimuli would require taking extra precautions to make sure that the animation processing does not slow down sampling too much / does not cause pauses in sampling.

Because NEWTRACK runs in DOS (e.g. FreeDOS [@FREEDOS]), which is a single-tasking operating system, there is no risk that other tasks would consume too much system resources at critical moments. DOS is *not* a real-time operating system, so all operating system services have no guaranteed timing properties. This causes no problems in our case, as operating system services (like reading data from a file) are only used between trials, not during trials.

Chronometer functions are implemented by querying the time stamp counter of the CPU: This 64-bit counter is incremented every clock cycle of the CPU. As a precaution, it is recommended to check the CPU data sheets if energy saving modes which freeze or slow down the counter exist, and if so, disable those in BIOS[1] setup. All Pentium compatible processors have a time stamp counter, and on the AMD Athlon 700 MHz processor, the counter ticks at 700 million ticks per second. NEWTRACK measures this frequency during initialization and uses it to establish a millisecond-precision time stamp base.

[1] Base Input Output System / firmware: bootstrapping software and basic drivers installed on a chip on the mainboard. Usually provides some operating system independent setup user interface.

### 2.1.4 Compact log files

Data about saccades (and eye *movements* in general) is not logged: Instead, only coordinates and times with *stable* gaze location – fixations – are logged. Together with information about buttonpresses and the properties of the displayed stimulus (e.g. image filenames and location of text on the screen), the amount of log data is very reasonable. A typical experiment session for our studies takes half an hour, half of which is spent for initial setup, participant briefing, recalibration of the system and similar tasks.

During the remaining time, at most a few 100 kilobytes of data are collected, which are kept in RAM and only stored to disk at the end of the session, to avoid delay and noises caused by disk access. All text is loaded at the beginning of the session, while images are loaded on a per-trial basis. If the noise or delay caused by this are to be avoided, the images can be copied to a RAMDISK[1] at the beginning of the session. The used image format is compressed PCX with up to 256 palette colors.

### 2.1.5 Working with NEWTRACK on various hardware

When you start NEWTRACK, you can provide the participant name / number (used to name the log file) and script file name (containing the stimulus definitions) either as command line arguments or interactively. You also have to select either text-only (default) or graphics (allows simultaneous image and text presentation) stimulus display mode through a command line argument. On systems with a modern “VESA VBE3” compliant graphics card for the participant display, you can also select a custom higher refresh rate (see [@SDD] and [@VBEHZ]) to reduce flicker.

On older hardware, NEWTRACK still supports plain text mode and can use the devices used in a typical old PCEXPT compatible PC. There are also fallback drivers for simulation purposes, which allow to test NEWTRACK and stimulus definitions for it on a normal PC without any eye-tracking hardware connected and without a second screen connected.

The standard version of NEWTRACK has several built-in drivers: Coordinates can be provided either by the custom NEWTRACK LTC1286 A/D box, by a conventional ISA DAS16 device (as used by the classic PCEXPT eye-tracking environment by Charles Clifton [@PCEXPT]), or, for simulation and testing purposes, by moving the “gaze coordinates” with a standard computer mouse (only enabled if no real coordinate source is active). If no CPU time stamp counter is found, a DJGPP library function is used as fall-back to provide high resolution time stamps.

Button presses can be taken from the A/D box or an ISA CIO device (usually on the same circuit board with a DAS16), but supporting joystick port connected buttons would only require small modifications in the sources. Button presses can always be simulated using the keyboard, but only a button box reaches the sampling rate / temporal resolution that justifies the use of millisecond-precision button press timestamps in the logs.

[1] A RAMDISK is a virtual disk created by a ramdisk driver out of RAM memory allocated from the operating system.

The second graphics card (used for the operator display) can be either a second PCI / AGP graphics card or a legacy monochrome graphics card (as used by PCEXPT). Note that all drivers for enabling NEWTRACK to work on old PCEXPT compatible computers are not well tested: One of the main design goals for NEWTRACK is to provide a solution which works on modern standard hardware. Classic PCEXPT, on the other hand, requires a PC with several now-exotic ISA components and makes no use of the hardware features of modern PCs. It is plain 16-bit software (you can compile it with the Turbo C compiler available for free in the Borland software museum [@BORLAND]) and has grown quite kludgy over the years. Minimum system requirements for NEWTRACK, on the other hand, are still very reasonable. Only a fraction of the 64 MB of RAM of our lab PC is used for the current experiments, and even using a slow (compared to office PCs today) 700 MHz CPU has not been a bottleneck so far.

### 2.1.6 Working with NEWTRACK and your existing software

NEWTRACK uses straightforward plain-text file formats for stimulus definition scripts and log files, so interfacing with your existing software should be easy: An example script is shown in the appendix A in table A.1, but only a small part of the supported scripting features were used for the concurrent scene and sentence presentation experiment. A full description of the stimulus script file format can be found in [Aue05].

Recent improvements in NEWTRACK include a READING compatibility mode, where scripts for the READING eye-tracking environment of our department are parsed directly, mapping a significant subset of the features to NEWTRACK features – usually enough to use READING scripts directly in NEWTRACK. Note that NEWTRACK and READING have completely different hardware requirements: The latter needs two computers, one running the image processing software under DOS, the other running the stimulus presentation software under Windows, using the SR EyeLink software library [@EYELINK] and SensoMotoric [@SMI] camera-based eye-tracking hardware.

The log files created by NEWTRACK use an easy to read and easy to parse plain text format which is explained in detail in the documentation ([Aue05]). A short example excerpt is shown below:

```
0000000 PARAMS hans set1\script6.txt 1280 800 96
...
0000000 TRIALID 5-b-5_1-OVS-m-b
0000000 INFO WORD 0 168 216 231 263 Die
...
0000000 INFO WORD 7 920 216 1079 263 Priester.
...
0000102 DISPLAY ON
...
0000443 FIXATION 198 234 703 260
...
0003487 ENDBUTTON 1
```

```
0003487 TRIAL OK
...
```

Each displayed word has the screen location logged, including half a space on each side and optionally some extra margin above and below each word). All fixations are logged with onset time, coordinates, offset time and duration.

### 2.1.7 The main menu architecture of NEWTRACK

NEWTRACK is designed to be a compact eye-tracking environment, so questions asked at program startup have been reduced to the minimum, and you can provide all information as command line arguments to start the experiment right away. After activating the hardware drivers, the raw coordinate data is displayed and the system waits for a button press to enter the initial calibration.

Calibration involves showing fixation targets to the participant sequentially, collecting raw coordinates for a $3 \times 3$ grid centered on the participant screen. After that, the conversion parameters for raw to screen coordinate transformation are updated, using a quadratic curve best fit. Given the quality of raw coordinates from the Fourward DPI eye-tracker [@WARD] and the quality of today's A/D converters, this has been good enough so far, but a possible enhancement would be showing more gaze targets to estimate parameters for a more advanced transformation function.

Conversion parameters are shown, and both the participant and the operator can see a gaze-controlled cursor while all nine grid marks are shown again simultaneously. This allows easy verification of the calibration quality: The operator asks the participant to look at some of the grid marks, while the operator display shows a gaze-controlled cursor. Mismatches between calibrated coordinate system and actual gaze location can easily be detected, the gaze cursor will not match the grid mark locations well in such cases.

The nine measured X and Y target coordinate pairs are used as a pool: For both X and Y, there are three possible screen coordinate values each. For each of those six values, three of the measurements contribute calibration data. From each such triple, the average of the two measurements which are closest to each other in result is used. There is no measure of calibration confidence yet: The third value of each triple is just ignored, there is no warning if the third value is too different from the other two or if all three values differ too much from each other.

After manually checking the calibration quality with the test grid, the operator can either restart the calibration again or start the first trial. Re-calibration mode uses a randomized display order of the gaze targets. The start or calibrate selection is done with the "next" and "calibrate" buttons of the operator button box, or, alternatively, with the enter and C keyboard keys. Starting the first trial enters the main trial loop, which is shown in table 2.1.

Each trial ends when either one of the button box buttons which was enabled in the stimulus / trial definition script file is pressed[1] by the participant or a timeout happens. The timeout

[1] Scripts can also define keyboard keys as valid response keys

delay is defined in the stimulus script as well, and recent additions to the scripting language, allowing various trial entry methods, allow the use of short timeouts to show priming images non-interactively as well. Using the keyboard, the operator can also abort trials (which restarts them) or abort the whole experiment. In either case, the software saves collected data before actually aborting.

## 2.2 Post-processing with DPIFilter

The data stored by NEWTRACK during our eye-tracking experiment required some post-processing prior to performing further analysis. For this, the DPIFilter software, an improved variant of AscFilter ([Kas03]) was used. DPIFilter consists of two parts: A system for data filtering and pooling and an interactive system for data visualization and corrections. The whole system is written in Java, which makes it platform-independent.[1]

### 2.2.1 The data filtering component

The data filtering and pooling component of DPIFilter loads a log file from NEWTRACK, converts it on the fly to the slightly different READING log file format, and performs several

Table 2.1: Overview of the course of events during a single trial. "LOG" marks events for which additional (millisecond-precision) timestamps are logged. The gaze location is shown on the operator screen in real-time, inspected regions are highlighted on the operator screen for real-time feedback.

| | | |
|---|---|---|
| 1 | | screen is **blank with a gaze mark ($\star$) in the middle** |
| 2 | | participant **fixates gaze mark**, operator can jump to calibration screen or continue to step 3 – trial restarts at step 1 after a recalibration |
| 3 | log | screen is blanked, stimulus image and text are loaded (takes up to 0.1 seconds on 700 MHz AMD Athlon with AGP graphics) |
| 4 | log | **stimulus is revealed** on next screen refresh ($\approx$ 88 refreshes per second for $1280 \times 1024$ resolution on 96 kHz screen) |
| 5 | | **participant looks at stimulus, fixations are logged** – each fixation is logged with time and coordinates, offline analysis assigns fixations to stimulus regions later |
| | | ⋮ |
| 6 | log | participant hits either **yes or no buttonbox button**, correctness of answer is checked offline later – participant is not told whether answer was right |
| 7 | | return to step 1 for next (trial or filler) stimulus |

[1] AscFilter is usually used on MacOS or Windows, while most of the DPIFilter sessions for the current study were done in Linux.

user-configurable filtering steps: Unnecessary log line types can be stripped, as can be the complete log data for filler items. The latter is done by showing a list of all trials found in the current log file and allowing per-trial and regular expression based selection of trials which should or should not be preserved.

The AscFilter version is able to select which of the eye channels[1] are used and what to do with logged blinks, saccades and pupil size information. NEWTRACK does not log the properties of the saccades between fixations, just the fixations themselves. The Fourward DPI eye-tracker only tracks one eye, without usually measuring pupil size. So the corresponding AscFilter features are not used in the DPIFilter version. The Fourward hardware is able to measure some pupil size effects optionally, but logging pupil sizes for fixations would require extra A/D conversion channels to get the data to the PC.

The ability to log blinks would require only forwarding a single binary signal from the DPI eye-tracker to the PC, so future NEWTRACK versions may support that. The version used for this experiment, however, uses a pure coordinate-based approach for fixation detection: Short blinks cause the DPI hardware to freeze the current coordinate, so they become part of the current fixation when NEWTRACK analyzes the coordinate movements. Longer blinks and other track loss conditions cause the DPI eyetracker to scan the coordinate range to recover the eye position. This causes a characteristic "motion" pattern which is easily recognized as not being part of any fixation.

### 2.2.2 The data pooling component

The filtering and pooling steps of DPIFilter cannot be used separately, but you can select no-effect configurations for them to simulate the effect of using one of the steps separately. Pooling is applied to fixations only, as NEWTRACK does not log blinks or saccades, nor were we interested in analyzing blink or saccade data.

Saccades shorter than a configurable threshold are not kept as a stand-alone event. However, if their coordinates are close to the ones of the fixation before or after them (at most 11 pixels Euclidean distance), they are merged with the fixations before or after them. Fixations which are still too short after all merging are removed from the logs.

Fixations which point to off-screen locations can be removed, but we did not use that function as NEWTRACK automatically clips those. Off-screen coordinates can be created by later coordinate system offset corrections with DPIFilter, but the software which analyzes the fixation data would not be troubled by off-scale coordinates either.

### 2.2.3 The visual components: Data display

DPIFilter provides a way to display the fixations for each trial, superimposed on a schematic view of the stimulus shown to the participants during the lab session which generated the logs. For this, text location information is extracted from the logs, and image location information

[1] READING can optionally log the data for both eyes simultaneously, as it uses an eye-tracker with three cameras, two for the eyes and one for localizing the display from the viewpoint of the user

is extracted from the stimulus description script. As NEWTRACK now also logs image location information, future versions of DPIFilter can draw their data purely from the log files, the only exception being the images themselves.

For the image data, a module derived from NEWTRACK is used: This module (DOS and Linux versions available) takes a PCX image file as input, analyzes the contents, and replaces all depicted items by filled areas. Each area is filled in an unique color which is selected based on the horizontal location of the item. For our current study, this means that we can automatically generate PCX images with, for instance, red blobs at the location of the left, green blobs at the location of the middle and blue blobs at the location of the right depicted character. Feeding all stimulus images through this filter generated a collection of "template images" for DPIFilter.

Figure 2.1 shows DPIFilter visualizing the stimulus for one of the experiment items: Labelled boxes show the location of the displayed text, including margins around the text. The actual text display uses no boxes and uses a fixed-spacing $16 \times 32$ pixel font, while DPIFilter just puts labels in the boxes using a default font. As you can see, the template image overlaps the text area. This happens because the text location is selected relative to the image contents, not to the total image size. Both image and text are horizontally centered. In our experiment, the vertical image location was fixed as 5 pixels from the bottom of the screen.

The buttons on the lower left of the interface allow the user to show or hide the fixations for the trial, either one by one or all at once. The "Style" button offers a toggle between just showing small circles at the fixation locations or showing circles labelled with fixation numbers and connected by lines to visualize the overall eye movement pattern, as shown in figure 2.2.

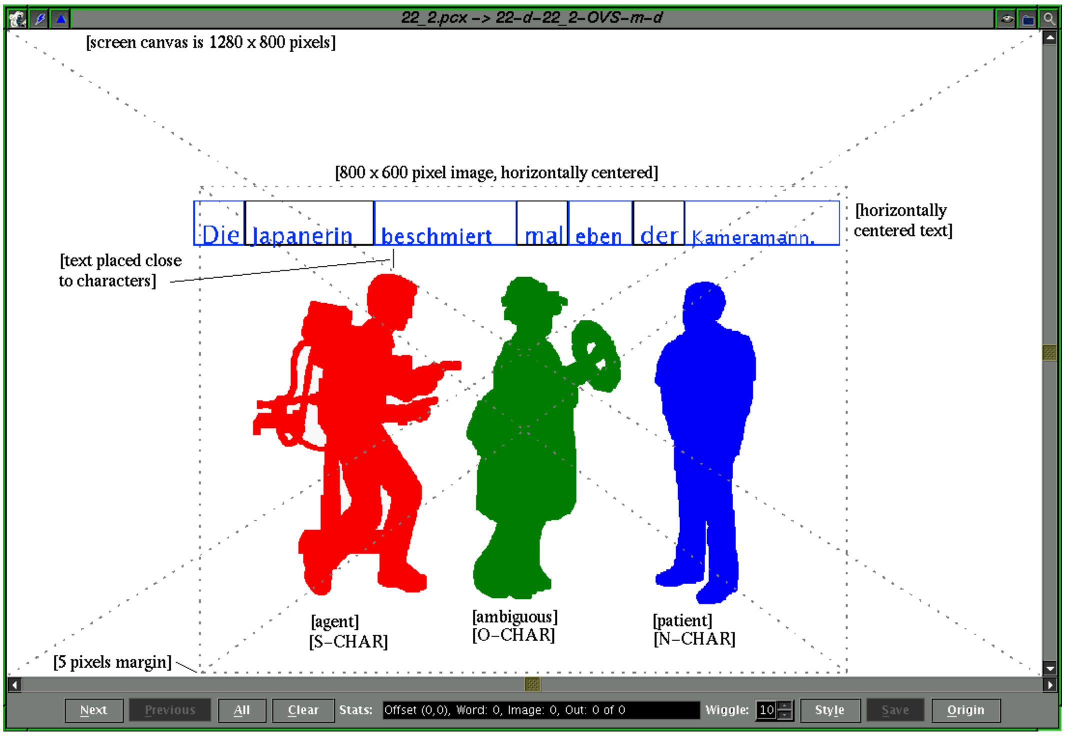
### 2.2.4 The visual components: Data correction

The second function of the visual interface is to allow the user to apply offset corrections to the data. This means that a coordinate system shift can be applied to all fixations of one trial simultaneously, by using the cursor keys or dragging the mouse. While the user updates the coordinate system, statistics about the number of assigned fixations (to text, image and background respectively) are shown in a constantly updated text field, which also changes color based on the percentage of background fixations. In addition, the correction vector itself is visualized as an arrow.

Fixations do not always fall exactly on an item. For text regions, this is handled by counting a certain margin around the words as belonging to the words themselves. For image items, however, surrounding all items with wide margins would make them overlap. Details of this problem are discussed in section 3.1.4.5. We call our solution with DPIFilter the "wiggle" function:

Wiggle means that for each background fixation candidate, pixels in eight directions (in 45 degree steps) are checked whether they are part of an image item, increasing the search radius in small steps until the user-selected maximum wiggle radius is reached. If this search hits an image item in unambiguous way, then the fixation is counted as being on that image item.

Figure 2.1: Layout of the stimulus display, as shown by the DPIFilter log viewing and log cleaning software. Colors adjusted for printing. DPIFilter uses a default font and shows boxes which include slack around the text. The actual display visible to the participants uses a fixed spacing (16 × 32 pixel) font and, of course, shows no boxes.



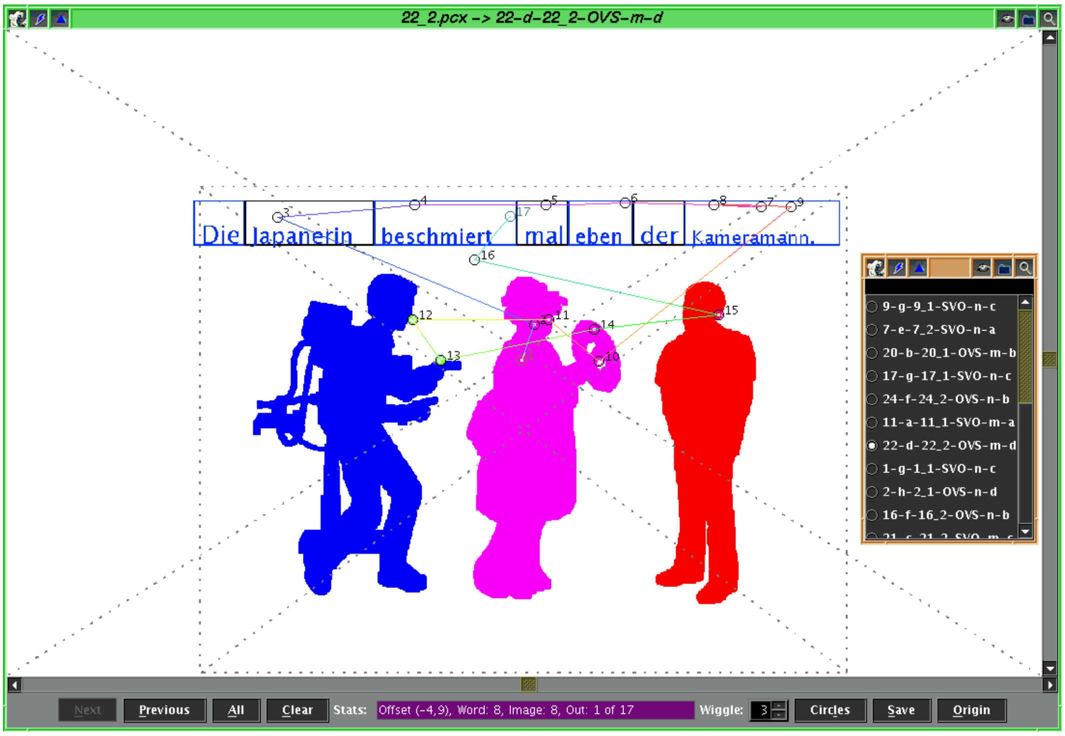

A stalemate situation leaves the fixation assigned to the background. Wiggle search does not consider text items as possible targets.

An example of using wiggle is shown in figure 2.2: Here, we moved the coordinate system offset by 4 pixels to the left and 9 pixels down to make most fixations fall on their target items. In addition, we set a wiggle radius of 3 pixels to allow fixations 12 and 13 to be recognized as being on the cameraman. Only fixation 16 is a true background fixation.

## 2.3 Summary

In this chapter, we have described the software used for our eye-tracking study. NEWTRACK got the main focus here, but DPIFilter has been a very helpful tool as well. The final generation of several dozen measures for each trial, which will be searched for patterns related to the condition (e.g. SVO versus OVS) in section 3.2, is done by a custom Perl script which is not described in detail here. The data analyis itself (not the script) is described later, in section 3.1.4.3, along with a description of the regions and measures.

Figure 2.2: Example of DPIFilter action (colors adjusted for printing), showing a trial and the fixations logged for that trial. The display of fixation numbers and connection lines can be toggled. In the screenshot, a 3 pixel "wiggle" is used to recognize fixations 12 and 13 as being on the cameraman. All fixations have been shifted by (-4,9) pixels to match the template better.




NEWTRACK is a new eye-tracking environment which can be used on standard PCs equipped with a few small additions: A second standard graphics card and some device (made of easily available standard components) to digitize the coordinates provided as analog voltages by (for example) a Fourward DPI eye-tracker. The software is easy to use and the stimulus definition and log file formats are designed for easy manual and automatic processing. NEWTRACK being open source software allows the user to add customizations like support for alternative hardware. The need for special hardware has been minimized by making use of the features of standard PC components. Some of those have not been used in any freely available eye-tracking solution yet. An example is the ability to use two PCI or newer graphics cards at the same time in DOS with a generic driver.

DPIFilter, an improved variant of AscFilter, is a Java tool which combines log-filtering and a graphical user interface to review fixation trails and apply coordinate offset corrections to logged fixation data manually. It helped to improve the usability of the logged data while the efficient user interface allowed to carefully manually correct the logged coordinates for all trials in reasonable time.

# Chapter 3

# Eye-tracking experiment

## 3.1 Method

### 3.1.1 Participants

A total of 32 students were paid 5 Euros (for half an hour in the lab) each for participating in our experiment.

Table 3.1 shows the age and gender distribution among the participants. Having only 28.1% male participants was not intended, there were simply far more female than male volunteers responding to our bid for helping science. Most participants were in their twenties, and there were 9.4% left-handed participants. As the Fourward DPI eye-tracker at our department can only track the right eye, no attempt was made to test for and track the preferred eye of the participants. Most participants were from other faculties: Many were students of psychology and pedagogics.

Some (15.6%) of the participants were from the department for computational linguistics, which the psycholinguistical department is a part of, but all participants were only allowed to take part if they had not been tested in studies employing the same or similar stimuli before.

Table 3.1: Age and gender distribution of the participants in the eye-tracking study. The call for participants was sent to a mailing list of participants in earlier psycholinguistics experiments and invitation posters were placed on advertising pillars and similar locations on the campus.

| Group | Count | | Age range | Average Age | Median age |
|---|---|---|---|---|---|
| All participants | 32 | | 19 ... 49 | 24.4 | 22.5 |
| Males | 9 | (28.1%) | 19 ... 27 | 23 | 22 |
| Females | 23 | (71.9%) | 20 ... 49 | 24.9 | 24 |

9.4% left-handed (all female), 15.6% students of computational linguistics.

### 3.1.2 Materials

#### 3.1.2.1 Stimulus design

As introduced in section 1.1.3, the experiment conducted for this thesis examines the effects of a match / mismatch manipulation and of depicted events on the comprehension of initially structurally ambiguous SVO / OVS sentences.

OVS word order is possible, but the SVO word order is preferred: [Hem93] found that OVS sentences take longer reading times, in particular the inspection of the subject NP takes longer for them. OVS word order is non-standard, and our sentences are designed to be initially ambiguous about word order until the case-marked determiner of the last noun phrase is reached. This case-marking makes it impossible to interpret the last noun phrase as sentence object. The initial ambiguity is possible by using female first noun phrases, for which the determiner is the same ("die") for both nominative and accusative noun phrases.

An example pair of sentences is: "Die Japanerin beschmiert (den[acc] / der[nom]) Kameramann." (The Japanese [woman] besmears the cameraman. / The cameraman besmears the Japanese [woman].) The noun phrase "The Japanese [woman]" can be both nominative or accusative, as the nominative and accusative forms of the female determiner are both "die". The actual experiment sentences were enhanced by an adverbial area, e.g. "Die Japanerin beschmiert soeben den Kameramann." (The Japanese [woman] besmears the cameraman right now.), to have more text around and to separate verb and second NP area from each other.

Figure 1.1 (see also figures 2.1, 2.2 and 3.4) shows an example image: There are three characters and two actions. Each stimulus text for that image mentions one of the actions and two of the depicted characters. From left to right, we see a cameraman, a Japanese and a policeman. The cameraman besmears the Japanese and the Japanese garlands the policeman. The image matches the OVS text shown in figure 3.4: "Die Japanerin beschmiert mal eben der Kameramann." (literally: The Japanese[acc/nom, female] besmears (just so) the cameraman[nom] / meaning: The cameraman (just so) besmears the Japanese.)

Had this been a "visual world" study, where sounds are played to the participants, the adverbial material would have given extra time for stimulus processing (and for eye-tracking during that processing) between hearing the verb and hearing the determiner of the second noun phrase. The action described by the verb is always shown somewhere on the image, but in mismatch case, it is performed by the "wrong" character. So the participants get the chance to first match verb and first NP with the image. It might surprise them that the first NP refers to the object of the action in OVS match case. Later, when reaching the second NP, they get the chance to check if the role relations established by the depicted action match the role relations expressed by the sentence. The adverbial region serves as a separator there. The adverbial region also serves as a separator in *our* experiment, because it prevents people from seeing verb and case-marked second NP determiner at the same time. In addition, the adverbial region acts as a buffer that prevents fixations right after the end of the verb or right before the second NP from being counted as fixations on the second NP or the verb respectively.

Each (SVO or OVS) sentence is shown concurrently with an image which either does or does not depict the action described in the sentence (match / mismatch). To preserve the initial sentence type ambiguity, the image always shows the character referred to by the first noun phrase in the middle, where it is both the agent of one action as well as the patient of another action, involving the characters shown on the left and on the right in the image. The direction of actions can be both to the left and to the right, and the characters on the left and on the right of the image are both male. This allows both of them to be named either by the subject or the object of the sentence (controlled by case marking of the sentence).

#### 3.1.2.2 Two contrasts combined – four conditions

Combining both factors – word order and congruence – results in having a balanced set of four conditions. This symmetrical approach allows us to check for effects of each of the two contrasts (and even the image direction balancing factor) separately. Having all factors balanced means that interferences can be eliminated for analysis: For example the number of match and mismatch cases shown to all participants of the study, as well as the number of image action to the left and image action to the right cases, are equal, cancelling out possible effects of matching and image action direction on the participants' way of reading sentences with either of the two word orders.

We expect that the match situation helps with comprehension, so there should be fewer fixations / shorter inspections on the stimulus: e.g. [UJR04] found this effect in their experiment. It is consistent with models like the [CJ75] one. In the results of [UJR04], the match / mismatch contrast affected the scene fixations more than it affected the text fixations.

Following [Hem93], we also expect longer inspection of adverbial and second noun phrase regions for OVS stimuli. The serial presentation experiment of [KC05] found effects of sentence type on *all* text regions, not only adverb and NP2.

If the findings of [KCSP05] are valid across modalities (their experiment used spoken utterances rather than written sentences), we expect to replicate an interaction effect of sentence type on the agent and patient regions: More fixations on the agent for OVS match trials and more fixations on the patient for SVO match trials (the [KCSP05] experiment used only match trials).

The experiment uses a Latin Square design, as explained for example at [@WOLFRAM]: Each participant gets to see an equal number of trials of each condition, and in total, an equal number of trials is shown for each condition. There are four conditions: All combinations of both (SVO or OVS) sentence types with both matching and nonmatching images. In addition, both images which show actions from left to right and images which show actions from right to left are used, for balancing, giving eight raw trial types used during trial list generation.

Each participant saw (after 6 training items in a separate run) 24 items mixed with 48 fillers in pseudo-random order, with one, two or three fillers between each pair of items. All possible variations of the items were distributed equally over all participants.

The four main conditions correspond to four different role assignments to the three characters shown on the image for a trial. The four sets of assignments are shown in table 3.2. Regularities in that table are:

The ambiguous character always turns out to be the one referred to by the subject of the sentence (S-CHAR) for SVO sentences. For OVS sentences, it is always the O-CHAR, the character which is named by the sentence object. The agent character is S-CHAR for SVO match cases. The patient character is O-CHAR for OVS match cases. For mismatch cases, agent is O-CHAR for SVO and patient is S-CHAR for OVS. There is always one of the three depicted characters which is *not* mentioned in the sentence at all: We call this character N-CHAR. A full example of the trial design (table A.2) and a full list of all experiment items (text and images, table A.3) can be found in the appendix A.

Each participant only gets to see one of the two variants of each experiment item image: Depicted actions are either left to right or right to left. Both the balancing factor of the action direction and the two factors (word order and congruence) of the conditions are distributed over the trial lists for the participants in a Latin Square way.

As for semantical and lexical bias, the materials have been checked in earlier experiments for using equal-frequency verbs and all-nonstereotypical actors – see [KCSP05]. For example there are experiment items showing a princess, a pirate and a fencer involved in the actions of washing and painting each other. It has been tested that people were able to tell which actions are shown on the images. However, stimuli are designed so that none of the actions (washing, painting) have stereotypical actors[1]. In contrast, a picture showing a detective, a thief and a millionaire, involved in the actions of capturing (detective catches thief) and stealing (thief steals from millionaire) would have caused a bias: People would have strong expectations about who does what to whom.

Table 3.2: Mapping of the image action roles (table head) to sentence roles (table body) for each trial condition. For image action direction ⇒, there is an agent on the left and a patient on the right. For ⇐, agent and patient are swapped.

| Word order | Do image and text match? | Agent character | Ambiguous character | Patient character |
|---|---|---|---|---|
| SVO | Yes | (N-CHAR) | S-CHAR | √O-CHAR |
| SVO | No | ⋆O-CHAR | S-CHAR | (N-CHAR) |
| OVS | Yes | √S-CHAR | O-CHAR | (N-CHAR) |
| OVS | No | (N-CHAR) | O-CHAR | ⋆S-CHAR |

[1] There have been no attempts yet to empirically measure the experienced stereotypicality of the actor / action combinations, though. It would also be an option to search for mentions of washing princesses and the like in a large corpus – frequency is assumed to be very low.

### 3.1.3 Procedure

#### 3.1.3.1 Lab equipment: DPI eye-tracker

This thesis describes an eye-tracking study about the interactions between seeing people doing things and seeing text describing those actions. The experiment uses classic equipment, a mostly standard PC, and some novel software.

The lab setup is shown in figure 3.1: The participant is sitting in front of a computer monitor which is showing text and images concurrently. To avoid head movements, the participant's head is leaned against forehead- and chin-rests mounted to a table. In front of the participant, there is a translucent mirror: Looking straight through the mirror, the participant can see the 21 inch (40 × 30 cm canvas) computer screen, 145 cm away. If the participant were to focus closer, he or she could see the reflection of some optical equipment in the mirror. As with looking through a window, the reflections do not disturb looking at the screen.

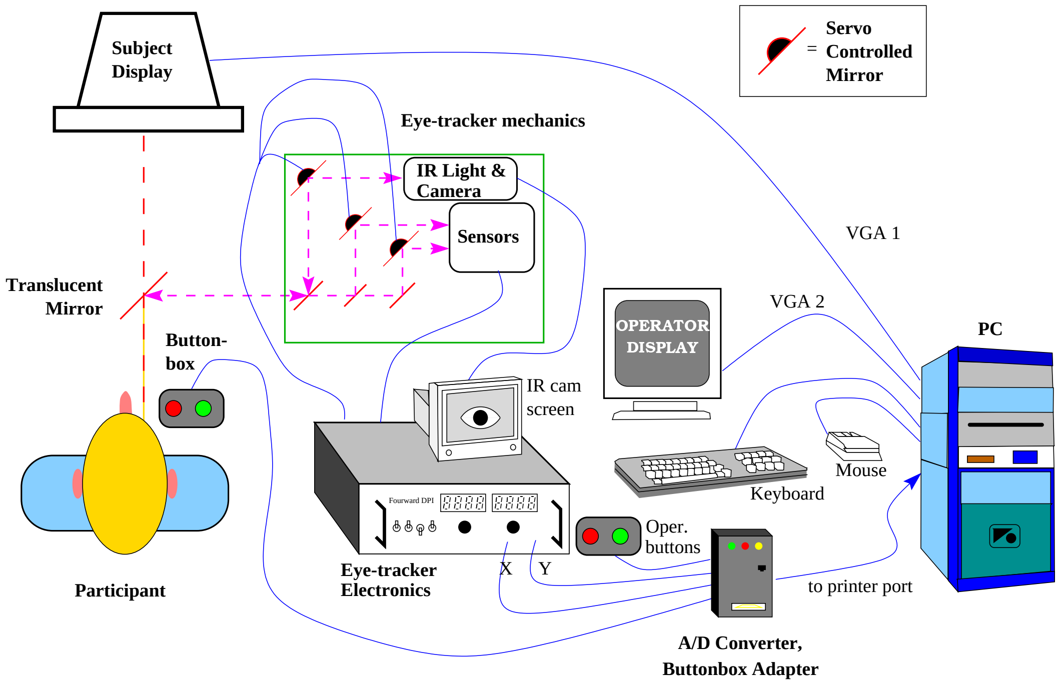
The optical equipment, however, is part of the eye-tracker, a machine which is "looking in your eyes" – optically gathering information about eye state. Optically measurable parameters include eye position, eye rotation, pupil size and lens shape / curvature. The DPI method is somewhat unstable – two small light reflexes in the eye have to be followed by a machine which is not mobile and not firmly attached to the participant's head. So the operator has to stay alert throughout the eye-tracking experiment, keeping meter readouts in range to help the eye-tracker to keep following the participant's eye movements.

Figure 3.1 shows the setup with a Fourward [@WARD] DPI (Dual Purkinje Image) eye-tracker. Other common eye-tracking methods use cameras, either head-mounted (attached to a headband) or located close to the screen. The latter usually requires additional mechanics to keep the cameras aimed at the participant's eyes. It reaches less precision, because the distance between camera and eye is bigger. On the other hand, the participant does not have to wear any sensor equipment, so the technology is useful for assistive systems for bodily disabled people. The head-mounted technology has the disadvantage of needing an additional camera to detect the screen location relative to the head. This is the main limiting factor for accuracy there, but large markers glued to the computer screen are still better to localize remotely than reflections in a person's eye viewed from the same distance from a camera located close to the screen. A general problem with camera-based eye-tracking is that the used cameras and the image processing needed to calculate the gaze position have limited speed: High-end eye-trackers like the fastest SMI system (see [@SMI]) can reach up to 250 or even 350 Hz sampling rate, but are far more expensive than a standard system used for assistive technologies.

Purkinje Image eye-trackers can work entirely without cameras: an invisible infrared light beam is aimed at the eye, using a servo controlled moving mirror. The reflection of this light beam is again passed through another moving mirror, constantly trying to keep the reflected light beam centered on a sensor. To know only the eye *position*, it would be enough to analyze the movements of the first mirror.

However, the eye position does not tell much about what you are looking at, even if you suppress most head movements. The eye rotation is far more important for that. The angle

Figure 3.1: Schematic view of the eye-tracking lab and equipment. The participant is sitting in front of a translucent mirror, chin and forehead leaned against supports, buttonbox in hand (and out of view, to avoid distractions).




between the light beam from the eye-tracker and the eye can be measured by tracking *two* beams of reflected light in parallel: One is the reflection of light at the outside of the cornea, the other is the reflection at the inside of the lens. This tracking style requires a third moving mirror. As with reflections on double glas, the distance between both reflected light beams depends on the angle between beam and reflective surface. This allows to find the angle by measuring the relative position of both reflections. Actually there are even four reflections: Both the lens and the cornea on both their inner and outer surfaces cause one reflection each. Following two of the reflections is enough to know the eye rotation, and it is best to use the outermost and innermost reflective surfaces for eye-tracking, given the anatomy of the eye.

There are camera-based eye-trackers which use the simpler approach of tracking the pupil location and the location of *one* of the reflections: This method is called pupile center corneal reflection (PCCR). Products based on PCCR like MyTobii by Tobii Sweden or LC Technologies Inc.'s Eyegaze systems allow to detect the gaze position of a free-moving user (just sitting or standing in front of the system) with help of a motorized camera, which can help people with disabilities a lot. Tracking only one reflection means lower requirements for the imaging resolution. However, speed and accuracy are usually quite limited: MyTobii reaches an accuracy of 0.5 degrees at a working distance of 50 to 70 cm, at 40 data points per second. Head movements compensation errors further limit accuracy. Eyegaze reaches 60 points per second at less than 0.7 cm RMS tracking error.

A DPI eye-tracker follows the innermost and the outermost reflections, those of the outer corneal surface and of the inner lens surface. So *two* movable mirrors to steer the reflected light on *two* sensors are needed. The eye-tracker electronics constantly update the mirror positions with help of high-speed servo motors, controlling a total of three mirrors for lighting, first reflection sensor and second reflection sensor respectively. The only camera in a DPI eye-tracker is an infrared camera to allow the operator to actually see the reflected invisible light shining in the participant's eye, to manually verify correct operation.

The electronics of the DPI eye-tracker do not process any images at all. Dozens of control loops (a Fourward DPI eye-tracker uses a total of several hundred op-amp circuits) are constantly busy keeping light beams at maximum intensity and their reflections centered on sensors which measure the intensity in their left, right, top and bottom quadrants separately. By combining the positions of all mirrors, the gaze direction of the participant can be calculated with very high accuracy.

As all processing is done in the analog domain, resolution, speed and accuracy are not limited by any of the "digital" problems of camera-based eye-trackers like refresh rate and image resolution. The latter limits in particular the accuracy of long-distance measurements like screen position (for head-mounted eye-trackers) or the corneal reflex position in PCCR systems. Instead, limiting factors for DPI eye-tracking are the movement speed of the mirrors and the ability to keep the light beams stably locked to the exact center of the sensors.

Using a DPI eye-tracker for our experiment has some advantages: All the processing is done by analog electronics, keeping computer resources free for other tasks, and the participant's eye movements can be analyzed with high temporal resolution. The timing resolution is high but finite: The frequency response of DPI eye-trackers is in the 100s of Hertzes, and our software can read the coordinates from the eye-tracker as often as it wants. The current Fourward DPI eyetrackers have a bandwidth of up to 400 Hz, see [@WARD].

#### 3.1.3.2 Lab equipment: a "normal" PC

The data processing part of the setup is explained in detail in section 2.1. In our case, all processing is done on a standard (700 MHz) PC equipped with a second PCI graphics card. Our graphics cards are one AGP (for the stimulus display) and one PCI (for the second display) VGA compatible graphics card. The second graphics card is connected to a display for the operator, showing real-time information about the experiment progress. The connection between hardware and software consists of a small NEWTRACK adapter box connected to the printer port of the PC. This box contains two monolithic standard A/D (analog/digital) converters, one for each of the two coordinates which are output by the eye-tracker as voltages. More details can be found in the software chapter and in [Aue05] – the latter also provides a circuit diagram.

This box also serves as the connection to two buttons on a box held by the participant, which are used for YES / NO responses. For convenience, the operator has two further buttons, labelled "next" and "calibrate", to step to the next trial (or through a menu in the software) and to enter calibration mode. Our operator buttons, like the participant buttons, have a mobile button box, which can be placed next to the keyboard or the eye-tracker or can be held by an operator at any other convenient place in the lab. We found that even when using flexible quality cables, bending and movement can wear out and damage cables over time. This can lead to hard to detect “loose contact” problems after several months of heavy use, in particular at the “box ends” of the cables.

#### 3.1.3.3 Preparations, calibration and trial triggering

Participants were forced to look at the center of the screen to start each trial, and the instructions (see section A.1 in the appendix) are designed to give no bias towards either checking the image or the text first.

After letting the participant read the instructions, six practice trials were run, to make sure that the participant understands the task and to get an initial calibration of the eye-tracker (the practice trials are shown as an example of the NEWTRACK scripting language in table A.1 in the appendix A). Then, the eye-tracker was paused and the participant could relax once more and ask questions about the procedure (but of course not about what the experiment was actually about). Some participants used the pause to reassure themselves which was the YES and which was the NO button, as the buttons were placed out of sight: The participants had to put the buttonbox on their lap during the experiment.

As the main experiment started with the second calibration of the eye-tracker, operator and participant already got used to the setup procedure for the particular participant, reducing later setup times and improving the calibration quality. Still, it was necessary to recalibrate from time to time. Section 4.1.1 shows how improving NEWTRACK can reduce the need for recalibration by using drift correction.

For analysis, the gaze positions were classified according to the screen layout description shown in figure 3.4. The actual stimuli always show the text above the depicted characters, with the vertical text location selected as between one or two line heights from the uppermost part of the shown scene (snapping to an invisible grid of text character locations). Both text and image were shown horizontally centered on a 21 inch computer screen with high flicker-free refresh rate. A computer view on the screen layout can be seen in figure 2.1. The analysis of image contents is done in a uniform, automated way by our software.

#### 3.1.3.4 Timeline of each trial

Each experiment trial starts with a blank screen with a gaze mark shown at the center (see table 2.1). As soon as the participant is ready to judge the next image / sentence pair, he or she fixates the mark. The operator can check if the software gets the right coordinates for the centered gaze position. If not, the operator can make adjustments to the eye-tracker hardware and / or enter the software calibration mode (where several gaze target marks are shown in random order, the participant has to fixate all of them, and new parameters for the mapping from raw to screen coordinates are calculated). As soon as the gaze position is stable on the center, the operator triggers the actual trial:

NEWTRACK blanks the screen, renders the stimulus screen, analyzes it (to be able to give real-time feedback about looked-at regions to the operator) and, a fraction of a second later, the stimulus screen is revealed, showing both text and image at the same instant. The participant will keep looking at the middle character for a moment (on average 0.19 seconds, as our data show) and can then move on by looking at other parts of the image or at the text.

### 3.1.4 Analysis

#### 3.1.4.1 Data pooling

The fixation pooling settings used for our experiment are: Saccades shorter than 60 milliseconds are not kept as a stand-alone event. However, if their coordinates are close to the ones of the fixation before or after them[1], they are merged with the fixations before or after them. Fixations which are still too short after all merging are removed from the logs.

Short fixations are generally considered not to be an information source for conscious linguistic processing: The used limit of 60 milliseconds is the same as in the experiment by Underwood et al. described in [UJR04]. Most encountered fixations were even a lot longer than that: For each region apart from "background" and N-CHAR, the average first fixation duration was at least 171 milliseconds. The longest average first fixation duration was the one for the "depicted character who is mentioned in the subject of the sentence", with 250 milliseconds. The longest average first fixation duration for a text region was for the second noun phrase: 226 milliseconds.

#### 3.1.4.2 Assessment of the reached calibration accuracy

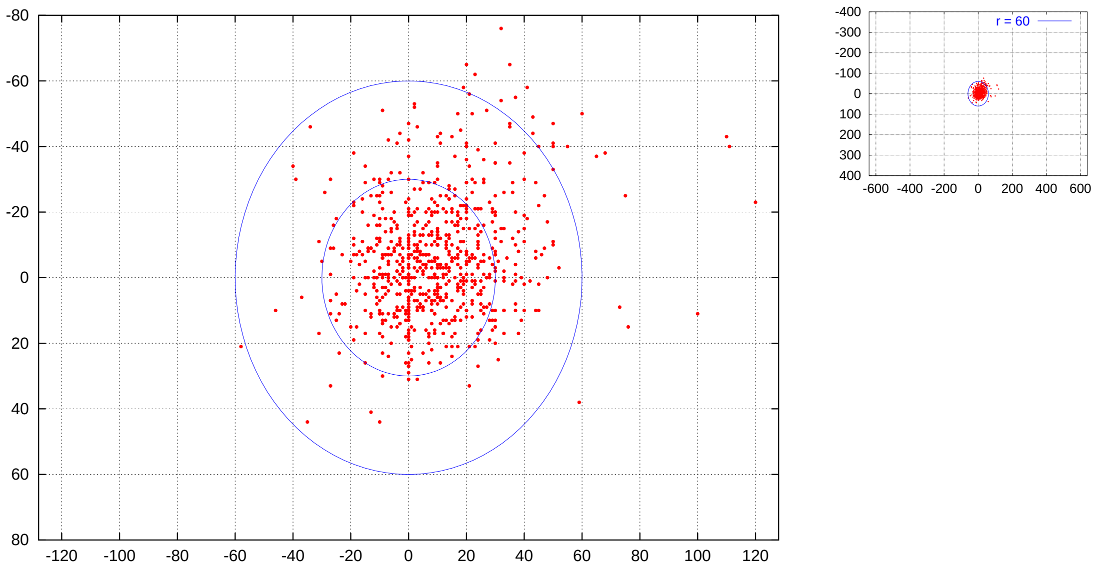
As explained in the software description, DPIFilter was used to apply corrections to the logged fixation coordinates, moving all fixation coordinates of a trial simultaneously to avoid data corruption. Figure 3.2 shows a scatterplot of the corrections applied to all trials except those with excessive track loss.

As can be seen, the vast majority of the corrections is at most one or two character's sizes (one character is $16 \times 32$ pixels big) away from the calibrated coordinate system of NEWTRACK. Still, this is enough to require this manual correction. Correction was done in a very careful and conservative way, for example trying to use similar correction vectors for consecutive trials. It would have helped to know between which trials recalibration had taken place. We could either modify NEWTRACK to log calibrations (including the coordinate system transformation parameters selected by them) or take notes in the lab manually to achieve that. The latter has the advantage of also capturing parameter adjustments to the hardware done without use of the NEWTRACK calibration screen.

Figure 3.3 shows an histogram view of the applied coordinate system offset corrections as well as the used wiggle values: The average correction vector was half a character to the right and one sixth of a character to the top. The former might indicate a systematic error

[1] at most 11 pixels Euclidean distance

Figure 3.2: Scatterplot of the coordinate offset corrections applied to the valid (track-loss below threshold) trials. Coordinates are in pixels, screen canvas size is (1280 × 800) pixels. The small plot to the right shows the corrections in the scale of the total screen canvas size, for comparison.



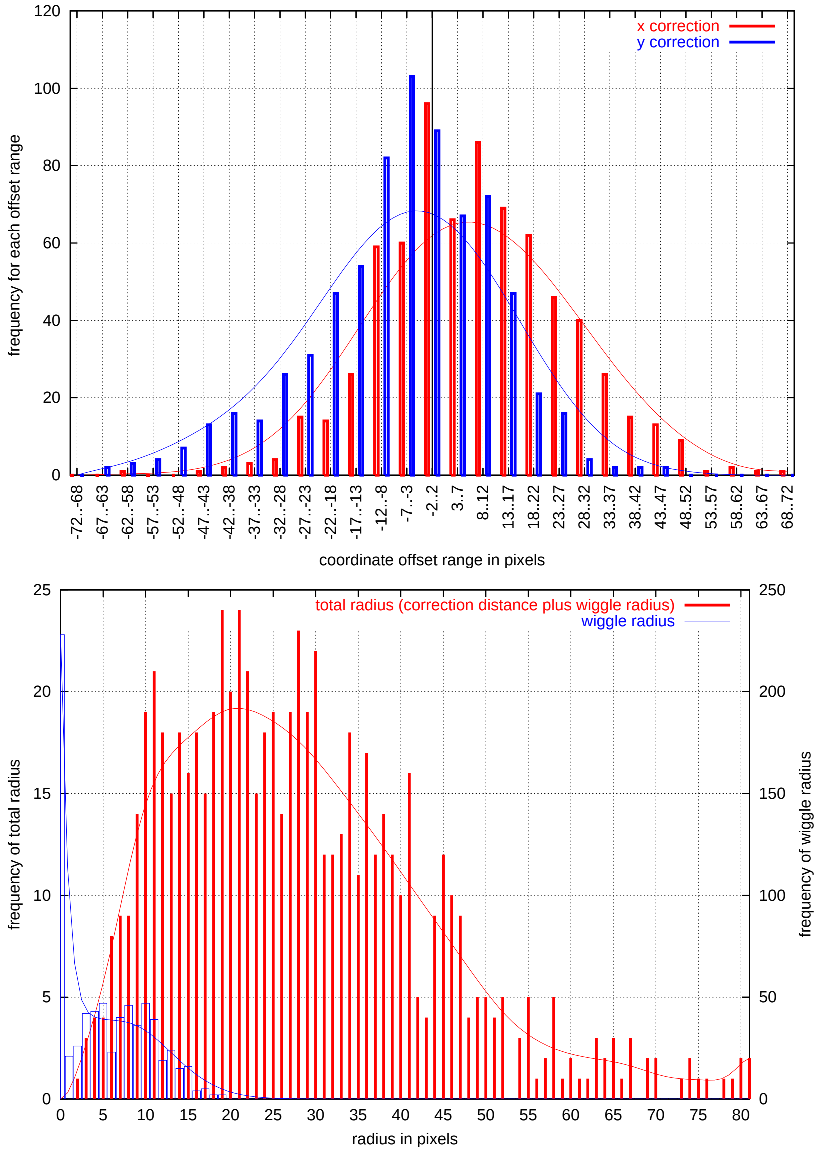

in the calibration system of NEWTRACK. A considerable improvement in coordinate offset accuracy is expected from the recently (after the experiment described in this thesis) added ability of drift correction in NEWTRACK, as explained in section 4.1.1.

The average wiggle search radius used in the manual corrections for this experiment was about 5 pixels, the average length of the correction vector was below 30 pixels. Many trials did not require any “wiggling” at all, while the others used wiggle radiuses well below 20 pixels.

#### 3.1.4.3 Distilling results from the fixation logs

From the DPIFilter software, we got a list of fixations for each trial, along with the button press information and the data and data about the stimulus properties like location of words on screen or condition (SVO / OVS, match / mismatch, along with the balancing factor: depicted action to the left / depicted action to the right). The final step to enable analysis of the results is to generate some well-defined measures from that data and search for significant effects of the conditions on those measures.

Figure 3.3: Histograms of coordinate offset corrections for the valid (track-loss below threshold) trials: Average correction vector: $(x, y) = (9.618, -6.062)$ pixels, average "wiggle radius" $r = 5.280$ pixels, average of $(r + \text{round}(\text{correctionvectorlength})) = 29.025$ pixels. Screen canvas size is $1280 \times 800$ pixels, stimulus font size (fixed spacing) is $16 \times 32$ pixels.



#### 3.1.4.4 The regions

At first, the stimulus display is subdivided into natural units: Depicted characters (including any items that they touch, e.g. tools used in the depicted actions), words of text, and the image background. To get more useful region definitions, the words of each sentence were clustered to groups corresponding to the four main components of the sentence: First noun phrase, verb, adverbial region and second noun phrase. This sectioning can be easily done because the noun phrases always consist of two words in this experiment, and the verb is always a single word, allowing automated mapping from word number to text region.

The image regions allow several viewpoints: First, the activity-based agent / ambiguous / patient role. Second, the mapping between sentence and image, which labels each of the three characters on the image as either S-CHAR, O-CHAR or N-CHAR (see section 3.1.2.2). Third, the location (left, middle or right) of the character. The middle character and the ambiguous character are the same. Items that touch a depicted character (e.g. depicted tools) count as being part of the region of that character.

In addition to the single regions, there are region groups: All text regions together form the region "sentence" and all image regions together form the region "image". The "anywhere" region is just the whole screen canvas, and the "background" region is all space outside sentence and image. Figure 3.4 gives an overview of the regions.

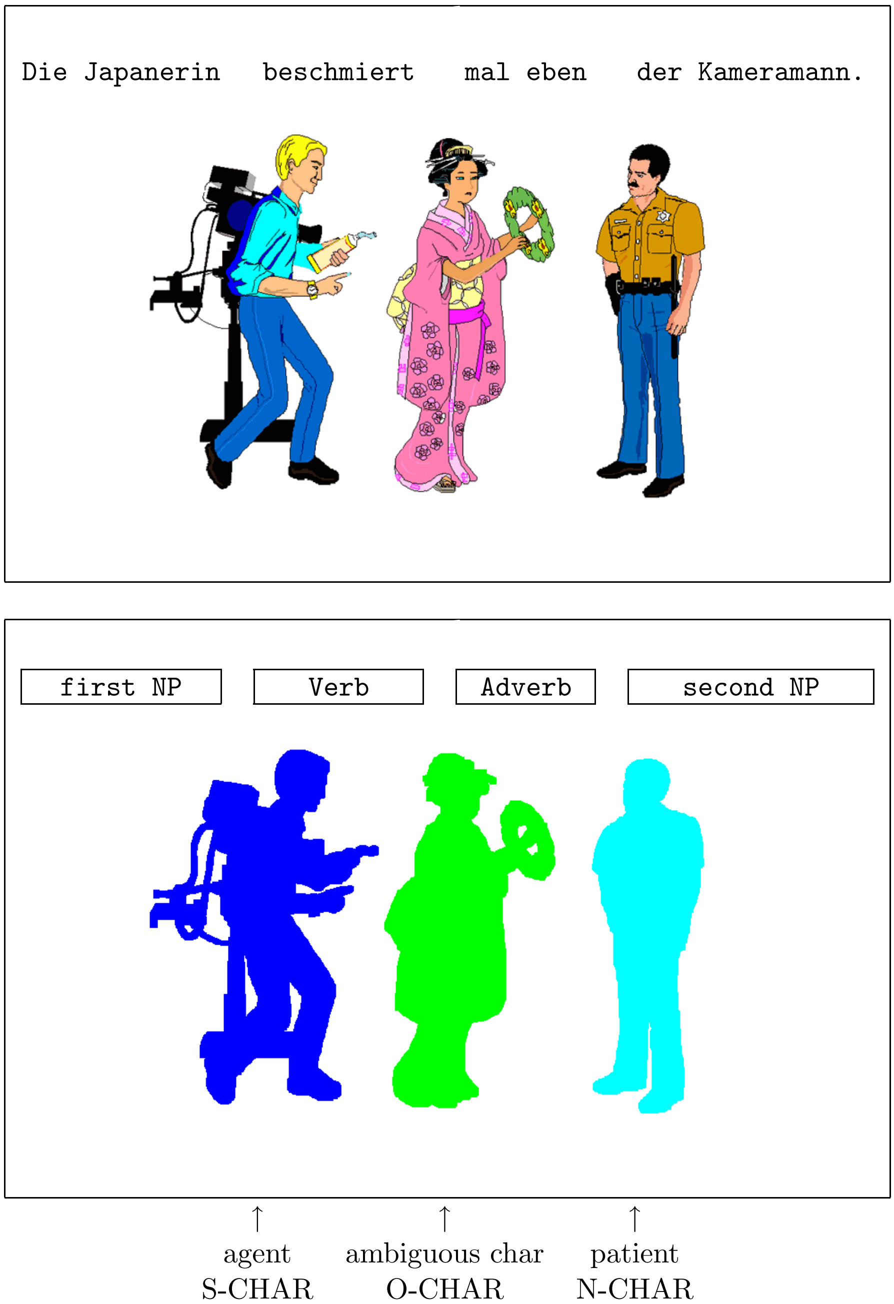
#### 3.1.4.5 Margins around regions

Text regions consist of the actual text (each character being $16 \times 32$ pixels big on a $1280 \times 800$ resolution screen) and half of the whitespace between the region and the next region to the left or right. Slack in vertical direction is a fixed amount of pixels above and below the actual text, in our case 12 pixels (less than half a row) each. A bigger value would be advisable for future experiments, as some of the participants read the text in a slightly diagonal gaze pattern. A good value might be half the character height above and below the text – in our case 16 pixels above and 16 pixels below.

For image regions, a border of 2 pixels around the actually visible character (and things touched by the character, like tools) was assumed to be part of the character. Further margin was added as needed, by letting the offline analysis software search for image items inside a "wiggle radius", which the real-time processing software does not do due to speed considerations. Making all borders around items more than 2 pixels wide would have spoiled the chunking process by making items touch each other.

As can be seen in figure 3.3, a search radius of 5 pixels on average (but rarely above 15 pixels) was enough to make most fixations "snap" to their items, leaving on average 5% of the fixations assigned to the background. Those were really on the background, and selecting a bigger search radius would bias the data / analysis. Yet again, a search radius of 14 to 18 pixels roughly corresponds to the "radius" of a character, so a search radius in this range is just as safe and advisable as the region slack for text regions recommended above.

Figure 3.4: Overview of location of stimulus regions in a sample trial: "image" is the combination of all three picture regions, "sentence" is the combination of all four text regions.



Using a constant value of 15 pixels would have similar effects to the manual choice, but the latter gives a better impression of how much "wiggle" is needed for this kind of experiment. Later experiments can use a fixed value. The software (DPIFilter) snaps to the closest item first, so even though adding 15 pixel margins around the items everywhere would make them overlap, no ambiguity in fixation to item assignment would be generated by always using a fixed 15 pixel search radius. However, computational cost grows with the search radius, which is also the reason why this kind of search and snap processing is to be done only in post-processing.

#### 3.1.4.6 The basic measures

As the initial fixation at the center image / ambiguous character is an induced one (participants had to look at the center of the screen to start a trial) this fixation has been removed from the measures which involve the ambiguous character in some way.

For each of the regions, a number of measures were generated based on the definitions in [SBC98]: The most standard pair of measures is the total number of fixations on a certain region during a trial and the total duration of those fixations. In addition, the duration of the first fixation on each region for a trial was taken as a third standard measure.

Apart from those straightforward measures, several "cluster" measures were taken. There were three types of fixation clusters, and for each of them, two measures are taken: The total duration of fixations in the cluster and the number of fixations in the cluster. However, at most two of the three cluster types were analyzed for each of the regions, as is illustrated in table 3.3.

For image regions, the only suitable cluster measure is the so-called first inspection: All fixations on the same region, starting with the first fixation on that region. So as soon as another region is fixated, the first inspection cluster has ended. It is not restarted when later fixations target that region again – those would be fixations in the second inspection cluster, or in the third inspection cluster, and so on.

#### 3.1.4.7 Some more complex measures for the text regions

For text regions, the first inspection measure was not taken. Instead, the more advanced measure of first pass was used[1]: A first pass fixation cluster is either empty or the same as the first inspection cluster. It is empty if and only if there have been fixations to words further towards the end of the sentence before the first inspection started.

The final most complex kind of evaluated fixation cluster is the regression path: The regression path for a text region *starts* when the first inspection to that region starts. So if there is no first inspection (because the region has been skipped already before it got looked at for the first time), then the regression path will also contain zero fixations for that region for that

[1]It would be possible to measure first inspection for regions for which only first pass was analyzed, but not the other way round: First inspection and regression path are only available for text regions with a defined left to right ordering.

Table 3.3: List of available measures for each region and region-group.

| | duration (time), number of fixations, duration of first fixation | first inspection **or** first pass: time and number of fixations | regression path duration and number of fixations |
|---|---|---|---|
| Region groups: | | | |
| anywhere | √, √, √ | — | — |
| background | √, √, √ | — | — |
| image | √, √, √ | — | — |
| sentence | √, √, √ | — | — |
| Image region classifications: | | | |
| left picture | √, √, √ | √, √ (1st insp.) | — |
| right picture | √, √, √ | √, √ (1st insp.) | — |
| agent | √, √, √ | √, √ (1st insp.) | — |
| ambig. char | √, √, √ | √, √ (1st insp.) | — |
| patient | √, √, √ | √, √ (1st insp.) | — |
| S-CHAR | √, √, √ | √, √ (1st insp.) | — |
| O-CHAR | √, √, √ | √, √ (1st insp.) | — |
| N-CHAR | √, √, √ | √, √ (1st insp.) | — |
| Text regions: | | | |
| first NP | √, √, √ | √, √ (first pass) | √, √ |
| verb | √, √, √ | √, √ (first pass) | √, √ |
| adverb | √, √, √ | √, √ (first pass) | √, √ |
| second NP | √, √, √ | √, √ (first pass) | √, √ |
| Other measures: | | | |
| button | Duration of stimulus presentation before YES or NO is pressed | | |

trial. The regression path *ends* as soon as a region further towards the end of the sentence is being looked at. This means that the regression path *includes* fixations to words *earlier* in the sentence as long as they are between beginning and end of the regression path cluster.

For both first pass and regression path calculations, fixations on image items were, like fixations on the background, ignored. There is no notion of any image item being further towards the beginning or the end of the sentence than one of the regions in the sentence. Image regions and the background are simply not part of the text.

## 3.2 Results

### 3.2.1 Analysis overview

A total of 32 participants were each shown 24 pairs of images and sentences. Each of the 192 possible combinations (24 sets of picture elements, used in two visual arrangements and combined with four variants of a sentence about them) was used four times. Each participant saw each of the 24 sets of picture elements exactly once. The design of the presented stimuli is discussed in detail in section 3.1.2.

All stimuli consist of an image, showing three characters involved in two actions. The actions are either both to the left or both to the right, both variants being used 50/50 to cancel out possible effects of the direction. We will, later in this section, have a look whether there were such effects, but the main purpose of using both directions is balancing.

The focus of our experiment is, however, on differences between SVO and OVS sentence types and on whether and how it matters if the sentence describes the image or if there is rather a clash between sentence and image. The experiment uses all four combinations of SVO / OVS and matches / does not match, which leads to the assignment between depicted characters and sentence elements shown in table 3.2.

A schematic view of a sample stimulus pair is shown in figure 3.4, the exact layout for another stimulus pair can be seen in screenshot of the log visualization component of DPIFilter in figures 2.1 and 2.2 in the chapter about software. The DPIFilter software allowed to pool or remove too short fixations, let us apply coordinate system offset corrections and, finally, assigned each fixation to a word, to an image item, or to the background.

For analysis, the stimulus display is sectioned into seven main regions (plus the background), four for the text and the three depicted characters from the left to the right. The text regions are, as described in section 3.1.2, the first NP (noun phrase), verb, adverbial region and the second NP, with only the latter disambiguating the sentence as OVS or SVO word order. The image regions were labelled in various ways: Activity-related (agent, ambiguous and patient) and sentence-related (S-CHAR, O-CHAR and N-CHAR). Those are considered to give more insights about how the eye movement pattern is controlled by the semantics of image, text or both, compared to the raw left / middle / right third labelling.

For each of the regions, in each of the labellings, the fixations and their durations were counted and clustered following each of several methods described in section 3.1.4.3. Some of the methods applied only to text regions, while others were only used for image regions. Table 3.3 shows that a large number of measures is used to describe the trials. Those measures are pooled by condition to get mean values and standard deviations for each measure for each condition.

### 3.2.2 Response button data

As first step, we separated the data into data for correctly answered and incorrectly answered trials: Only 74% of the trials got the right YES / NO decision from the participants. Closer inspection shows that while only 8.5% of the answers are wrong for SVO type sentences, 43.5% of the answers are wrong for the OVS sentences. Far too many to just throw away, so we decided to create a separate analysis of all the incorrectly answered trials. Be warned that, as there are only few SVO incorrectly answered trials, data is relatively sparse for that condition. An overview of the right and wrong answer counts is given in table 3.4. 47 trials had to be removed because they contained excessive durations of track loss.

Wrong answers are distributed quite evenly between the two possible image action directions and between the match case and the mismatch case. There is some slight tendency to give the NO answer for match sentences more often than to give the YES answer for mismatch sentences, though. A look at the raw data (see table 3.4) showed that this tendency does *not* come from the OVS items. Instead, they come from people being pessimistic about SVO match sentences. Being part of a linguistic experiment probably made the participants a bit oversensitive about small mismatches, e.g. about whether things are "obviously" or "right now" (two of the common adverbial region texts).

### 3.2.3 Trials lost due to track loss

Apart from having to apply some manual corrections to the coordinate system offsets of each trial, there were also problems with track loss conditions: The eye-tracker was unable to follow the participant's eye movements for an excessive amount of time during some trials.

Excessive track loss is defined, for our purposes, as 25% or more of the trial duration being gaps of more than 250 milliseconds between fixations (or between the last fixation and the final button press, that is). It turned out that 47 of the presented 768 trials are lost due to excessive track loss, only one of them explicitly marked as aborted during the experiment. A few trials were also repeated during the experiment, but doing multiple repetitions of the same trial reduces the information value of the logged data, as the participant has seen the stimulus several times. Table 3.4 shows the distribution of lost trials over the trial conditions.

Track loss is not a problem caused by software. Rather, it happens when control loop parameters of the eye-tracking hardware go out of range, for example if pupil size changes too quickly or eyelashes obstruct the optical path of the tracking system. It is also the case that the Fourward DPI eye-tracker has only limited compatibility with contact lenses. Those form extra reflective surfaces which can lead to incorrect measurements.

The lost trials were spread quite unevenly over the participants, as can be seen in table 3.5 discussed in section 3.2.5: The worst-trackable three participants had 10, 9 and 6 lost trials out of 24, respectively (not counting lost fillers, as no fixation data was logged for those anyway). All other participants had at most 3 lost trials each, on average one lost trial per participant.

### 3.2.4 Incorrectly answered trials

Another issue is that OVS word order sentences elicted a whole lot of incorrectly answered trials, only about half of all OVS trials being answered correctly. While this may, at first glance, give the impression that the participants were just guessing the YES / NO decision for OVS trials, there were actually two big groups of participants with different strategies: One group got OVS sentences right (still with more frequent erring than for the SVO sentences), while the other group just treated the OVS sentences as if they were SVO sentences, denying the case marking of the second NP that had to force them into reading the sentence as OVS. A third group of participants did indeed err quite often on OVS sentences, but it did not seem to go as far as just guessing their answers.

Table 3.4: Distribution of trial results over the conditions: Each of the 32 participants was shown 24 trials and 48 filler items. Trials with excessive track loss (more than 25% of the trial duration) were removed.

| 1. Trial type: | match | | mismatch | | |
|---|---|---|---|---|---|
| | SVO | OVS | SVO | OVS | Sum |
| Right answer: | 76 / 83 | 49 / 51 | 85 / 90 | 51 / 50 | 535 |
| **Wrong** answer: | 12 / 9 | 40 / 38 | 6 / 4 | 39 / 38 | 186 |
| Removed: | 8 / 4 | 7 / 7 | 5 / 2 | 6 / 8 | 47 |
| Shown: | 96 / 96 | 96 / 96 | 96 / 96 | 96 / 96 | 768 |

Each column shows data for ⇐ / ⇒

| 2. Trial group: | sentence type | | congruence | | balancing | |
|---|---|---|---|---|---|---|
| | SVO | OVS | match | mismatch | ⇐ | ⇒ |
| Right answer: | 334 | 201 | 259 | 276 | 261 | 274 |
| **Wrong** answer: | 31 | 155 | 99 | 87 | 97 | 89 |
| Removed: | 19 | 28 | 26 | 21 | 26 | 21 |
| Shown: | 384 | 384 | 384 | 384 | 384 | 384 |

### 3.2.5 Errors listed by participant

A final interesting thing to look at before getting to the detailed log data is which participants had which problems with the experiment. Table 3.5 gives a compact overview of the encountered troubles. For space reasons, up to 3 participants are listed in each table row. For example, the third row tells that participant 5 gave three wrong answers for OVS mismatch trials (i.e. this participant pressed the YES button while sentence and image actually did not match), four for OVS match trials and one for SVO mismatch trials. Participant 1 gave only 5 wrong answers for SVO match trials. So participant number 1 has the strong opinion that OVS sentences are never right / matching. Participant 22, on the other hand, and (weaker) also participant 32 were overoptimistic about matching in OVS case. They might have just assumed that there will be a match without taking the time to understand the OVS sentence fully.

Only one participant (number 27) made more errors for SVO sentences than for OVS sentences – one of the two participants which did all OVS trials perfectly. A total of 13 participants decided correctly about match / mismatch for all SVO trials, the others had a quite low error rate for SVO as well. You can also see a slight bias towards getting SVO match trials wrong more often than SVO mismatch trials for maybe three of the participants.

Many participants made a lot of errors for OVS sentences and 10 of them even performed a lot worse than if they had just been guessing the answers. A possible explanation for this is that those participants rejected the OVS sentence construction completely, in particular as – the first NP being ambiguous – only the case marking of the second NP gives them evidence that the sentence is not a normal SVO one.

Table 3.5: Distribution of problems over participants: Numbers of **wrong** answers and (due to excessive track loss) removed trials for all participants. Removed trials and trials with **wrong** answer can overlap. Compact listing with up to 3 participants in each table row.

| Participant numbers | **Wrong** answers for: SVO match | SVO mismatch | OVS match | OVS mismatch | Removed trials |
|---|---|---|---|---|---|
| 6, 7, 8 | 1, 0, 1 | 0, 0, 1 | 0, 2, 1 | 1, 0, 1 | 0, 1, 0 |
| 4, 2, 3 | 0, 0, 0 | 0, 1, 0 | **6**, **5**, **6** | **5**, **5**, **5** | 3, 0, 3 |
| 5, 1 | 0, 0 | 1, 0 | **4**, **5** | 3, 0 | **9**, **6** |
| 13, 12, 15 | 0, 0, 1 | 1, 0, 0 | 0, 1, 3 | 1, 0, 2 | **10**, 1, 0 |
| 9, 16, 10 | 1, 0, 0 | 1, 0, 0 | 1, 1, 1 | 1, 2, 0 | 1, 0, 1 |
| 14, 11 | 1, 0 | 0, 0 | **6**, **4** | **5**, **6** | 0, 2 |
| 23, 24, 22 | 0, 2, 1 | 0, 0, 0 | 1, 0, 0 | 0, 2, **5** | 0, 0, 1 |
| 21, 17, 20 | 0, 0, 1 | 1, 0, 0 | **6**, **5**, **6** | **4**, **6**, **6** | 0, 0, 2 |
| 19, 18 | 2, 3 | 0, 0 | **5**, **6** | 3, **6** | 0, 0 |
| 31, 26, 27 | 0, 1, 2 | 0, 0, 1 | 3, 3, 0 | 0, 2, 0 | 0, 1, 3 |
| 28, 30, 29 | 0, 1, 0 | 0, 2, 0 | 3, 1, 0 | 3, 3, 0 | 0, 0, 3 |
| 32, 25 | 1, 0 | 0, 1 | 2, 0 | **5**, 2 | 0, 0 |

Table 3.5 lists *all* trials: Even trials which got discarded later because of excessive track loss are counted in the columns which show the number of wrong answer. This makes sense because only the fixation data is unusable for those trials, while the button selection is, to some degree, useable. For all other analysis, not even the button press data is used for discarded trials: Participants noticing the eye-tracker problems (or being asked to blink to help the eye-tracker lock again, for example) possibly answered slower.

Table 3.5 shows that most of the discarded trials are linked to only three participants: Numbers 5, 1 and 13. For all other participants, at most three out of 24 trials had to be discarded because of tracking problems, the average among those 29 participants being less than one lost trial per participant. Track loss problems are commonly related to things like eyelashes getting into the optical path, lighting changes causing strong pupil size changes, or bad Purkinje image contrast.

After the lab session, three of the participants (14, 21 and 20) declared that the determiners (nominative case markers) of the last NP of some sentences contained spelling errors or that some sentences had generally wrong syntax. Those are, not surprisingly, three of the ten participants who had treated OVS sentences as if they were SVO sentences.

Five other participants had noticed that the experiment did something with case, that there was a striking amount of “der” and “den” determiners, that the subject location was relevant or that they often had to check the last determiner twice. For three further participants, some sentences just looked odd or complicated – one of them just answered YES for most OVS sentences. One participant, 28 who had guessing-like correctness on OVS, explicitly mentioned having noticed that the images all show three characters[1]. Finally, one participant said that the (depicted) character was "sometimes standing on the wrong side".

### 3.2.6 Descriptive presentation of the data

#### 3.2.6.1 The overall histogram view

Figure 3.5 shows a fixation histogram of all correctly answered trials. Histograms later in this chapter and in the appendix also show only data for correctly answered trials, but are further subdivided by trial condition. For example there will be one histogram with all (correctly answered) SVO trials and one with all (correctly answered) OVS trials. Figure 3.5 shows data for all regions: The Histograms later in this chapter only show data for a selection of regions and only for a narrower range (3rd to 25th) of fixation slots of the trials. You can find extended versions (showing data for all regions) of those histograms in the appendix B.

The histograms show percentages of fixations, not absolute numbers. For example the figure 3.5 shows that a bit more than 40% of the 5th fixations (of all trials which had 5th fixations at all) are on the verb region. Actually almost 99% of all correctly answered trials had at least 9 fixations.

More than 90% of all correctly answered trials had at least 11 fixations. More than 75% of the correctly answered SVO / OVS trials had at least 13 / 15 fixations respectively. Almost 50% even had at least 17 / 20 fixations for SVO / OVS respectively. More than 20% of all trials contribute data even for the last fixation slot shown in our histograms (the 25th slot).

As the participants have to fixate the center of the screen to start a trial, more than 90% of the first fixations of all trials are on the center image character. The second fixation of 40% of the trials is already on the first NP, but for another almost 30% of the trials, it is on the background (between the left depicted character and the first NP usually). This could be an artifact of the participants underestimating the center / vertical text location distance when planning their first saccade. As saccades are "ballistic", their target cannot be corrected in "mid-flight". Some 10% of the second fixations are on the verb instead of being on the first NP, which could again be an artifact of too short initial saccades.

Another possible explanation for the probability distribution in the second fixation slot is that it relates to the so-called "attentional blink": In a quick succession of stimuli, like animated text display, about 0.2 to 0.5 seconds after the onset of the first stimulus, there is a period of decreased attention (0.3 to 1 second, depending on the age of the participant).

It has been hypothesized that the brain is busy with adjusting the vision to cognition filtering process during that time (visual processing can handle $\approx$ 2000, but conscious visual processing can handle only $\approx$ 40 bits per second), leading to the loss of some of the seen information. While we have no rapid serial presentation here, our experiment still involves the sudden appearance of the concurrent scene and sentence display. The second fixation slot contains an unusually high amount of background fixations and short fixations (less than 100 milliseconds). The timespan between stimulus onset and the second fixation is usually in the same range as

[1] Which is not actually true for the fillers

the onset of the attentional blinks demonstrated in rapid serial visual presentation (RVSP) experiments. See [VZBdL99] for an example measurement of the attentional blink and for some additional pointers to further background information.

More than 80% of the third fixations are to the first NP. So there is a clear preference to read the sentence first before inspecting the image further. This is consistent with other experiments: For example [RRS+01] found that, for viewing print advertisements, participants started to read the text on average on their third fixation (first fixation is an induced one at the center of the display as with our experiment). They only returned to inspecting the image closer after having read a fair amount of the advertisement text.

Probability peaks for NP1, verb and adverbial region are at the 3rd to 4th, 5th and 6th fixations, respectively. The first peak of NP2 fixations is around the 8th / 9th fixation slot. So the histogram shows a typical left to right reading pattern over all four text regions. Peak areas get more stretched along the time (fixation number) axis for later peaks, as more variation accumulates from the earlier-visited regions.

Around the 10th fixation, half of the fixations are already on the image, the other half are still on the text. We can see that there are far more fixations on the middle image than on the other images throughout the whole trial. The middle character is always referred to by the first NP of the stimulus text in our design. There are also slightly more fixations on the image agent (the character who would be S-CHAR with an OVS sentence for which the middle ambiguous character would be O-CHAR) than on the image patient. With a SVO sentence which mentions the ambiguous character as S-CHAR, the patient character would be O-CHAR.

In general, the agent region gets more fixations than the patient region and S-CHAR gets considerably more fixations than the O-CHAR region. A reason for this could be a combination of a general bias to look at active / agent characters and a strategy to look at the S-CHAR to compare it's actions with the sentence subject ones. The first peaks for S-CHAR and O-CHAR are at the 12th and 14th slot, respectively. The S-CHAR and middle image data are quite similar, possibly because most of the correctly answered trials are for SVO cases where S-CHAR and middle image are the same. Differences will become visible when we look at SVO and OVS cases separately.

Data gets increasingly blurred across fixation slots towards the right of the histogram: The trial durations differ, so for example the 10th fixation slot can combine information about the last fixation of short trials and information about an earlier phase of longer trials.

It can still be seen that there is a weak "second wave" of higher amounts of text region fixations: The NP1 region gets more fixations again after the 12th slot, with a peak around the 16th fixation slot. The other text regions (again in their left to right order) have similar peaks at later slots. This could point to an eye movement pattern where some participants re-read the whole sentence around that time. The second peak for NP2, around the 21st slot, is stronger than those for the other text regions, which could point to another eye movement pattern where only NP2 is re-read. The probability for a fixation to be on the sentence type disambiguating second NP is always higher than the probability for a fixation to be on any other text region after the 6th fixation of each trial.

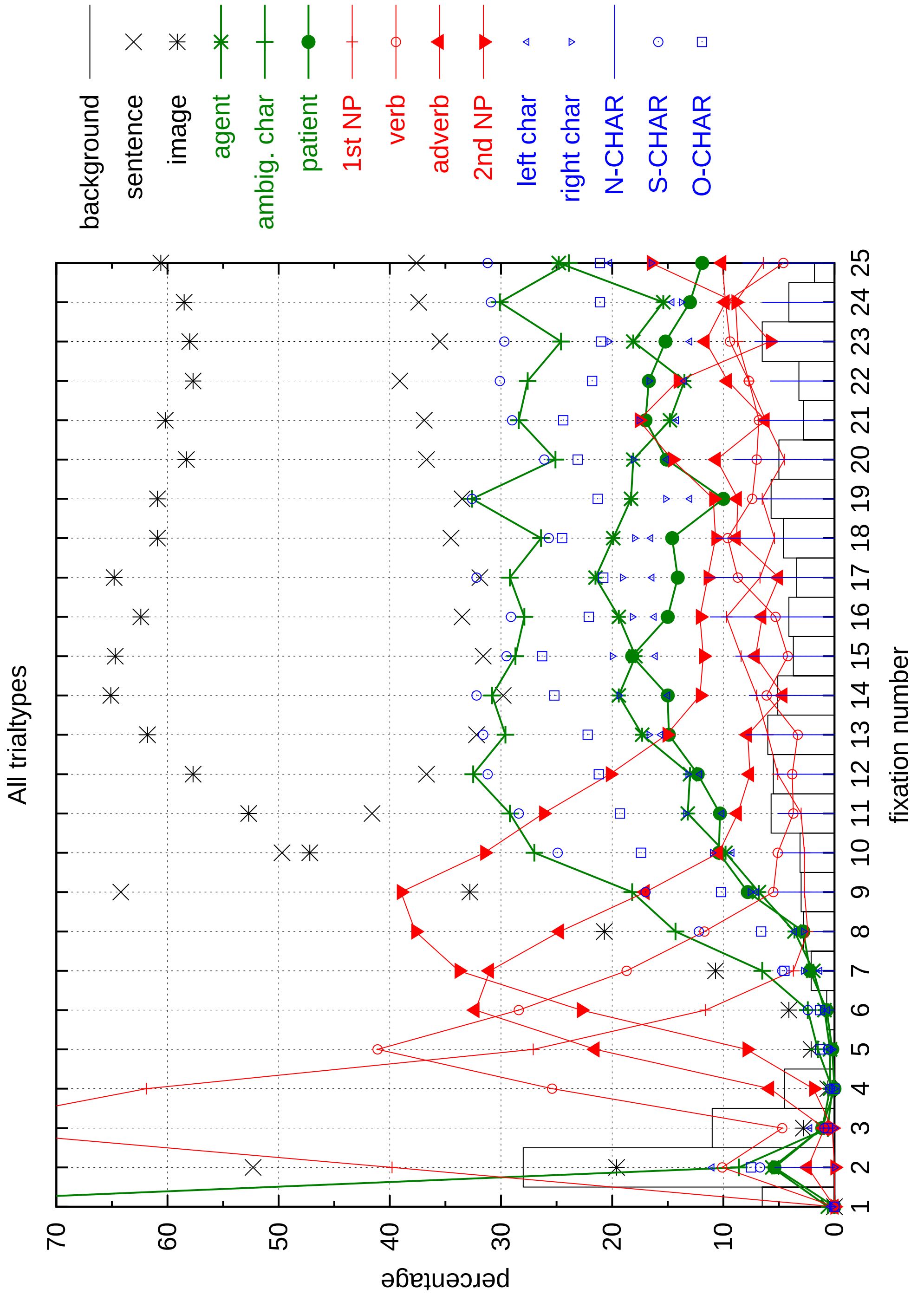



Figure 3.5: Full histogram of fixation frequency by fixation number: Totals for all regions for all correctly answered trials at all.

The probability for a fixation to be on the background (data drawn as boxes in the histogram) is only about 5% for any slot after the third fixation one. The probability for a fixation to be on the depicted character who is *not* mentioned in the sentence (N-CHAR – data drawn as impulses in the histogram) is also well under 10% for all slots except for the 16th to 18th ones.

We also see some differences in fixations to the left and right character. Data for those regions is shown as small up and down triangles in the plot. As the diagram combines data for trials for both image action directions, the small remaining differences in fixation distribution between both regions could show general viewing preferences: Starting at about the 11th fixation slot, probability for any fixation to be on the right image is slightly higer than the corresponding probability for the left image. The amount of this difference is quite small except for the 14th to 21th slot.

Next we move on to describing contrasts between the four conditions (SVO match, SVO mismatch, OVS match, OVS mismatch) in histogram data. We start by comparing the histograms specific to SVO and to OVS sentences, respectively, using reduced histograms. Those show only the data for the S-CHAR, O-CHAR and N-CHAR, agent, patient and middle image regions, the data for the text region of the second NP, and the data for the "any image" region group. The verbose versions – showing data for all regions – of all histograms used in the rest of this chapter can be found in the appendix.

#### 3.2.6.2 SVO versus OVS case

Comparing SVO to OVS cases (figure 3.6), differences start around the seventh fixation: Probabilities that fixations are on the second NP are (slightly) higher for the OVS case. More importantly, only 10% of the fixations for the 14th to 18th slot are still on NP2 for the SVO case, while this percentage stays up at 15% until the 16th slot for OVS. This could be interpreted as OVS causing more attention to the NP2-subject, compared to the more "normal" case of SVO NP2-object.

There are also far less image fixations for OVS trials in general: The chance that a fixation is on any of the picture regions at any fixation slot is at most 60% for OVS but up to around 70% (at 14th to 17th and at 24th slot) for SVO. Conversely, this means that there are more text fixations for OVS than for SVO at those slots.

SVO trials feature a considerably higher amount of fixations (compared to OVS cases) on the middle ambiguous character (data drawn as boxes in the histogram) which is named by the first NP. So there are more fixations to the middle character when it is the S-CHAR as well (compared to when it is the O-CHAR). For the 7th to 25th slot, S-CHAR gets more fixations than O-CHAR. First S-CHAR peak at the 12th slot is before the first O-CHAR peak at the 15th slot.

For OVS sentences, there is no big difference between the fixation probabilities for the S-CHAR region and the middle (here: O-CHAR) region in most slots. Now O-CHAR gets more fixations than S-CHAR for the 7th to 12th and 18th slot, while it is the other way round for the 13th to 17th and 20th to 25th slot. The first O-CHAR peak at the 10th / 12th slot is before the first S-CHAR peak at the 14th slot. In summary, there are more fixations on S-CHAR than on O-CHAR for both SVO and OVS, but OVS has more O-CHAR fixations than S-CHAR fixations in early slots.

The amount of fixations on the agent picture is considerably higher than the amount of fixations on the patient picture, but *only* for the OVS trials.

For SVO, the difference in fixation counts for S-CHAR (mentioned by the first NP for SVO) and O-CHAR is big, while the difference for agent / patient is very small. For OVS, it is *the other way round*: The difference for S-CHAR / O-CHAR regions is small, while the difference for agent / patient is big. The OVS agent / patient differences start in the 8th slot while the SVO S-CHAR / O-CHAR differences already start at the 6th slot. As mentioned earlier, OVS has S-CHAR / O-CHAR differences like SVO at first (from 7th slot on) but those differences *change direction* after the 12th slot for OVS.

One surprising pattern is that the probability for fixations on the N-CHAR is often higher in the SVO case, in particular around the 17th (15th to 19th) fixation slot.

#### 3.2.6.3 Match versus mismatch case

Figure 3.7 shows (for some regions) the data for the trials where image and sentence matched and the trials where they did not: The percentage of fixations on the middle character (which is named by the first NP / is "NP1-CHAR") is similar for both cases, except for a somewhat higher percentage in the 19th fixation slot for the match case. However, in the match case, there are considerably more fixations on the S-CHAR than in the mismatch case. Remember that in the mismatch case, the S-CHAR shows a character doing something else than the action described in the sentence. Rather, some other depicted character does the described action *with / to* the S-CHAR character as the patient of that action.

In the mismatch case, differences in fixation percentages for S-CHAR and O-CHAR regions are smaller after the 10th slot (compared to the match case). However, there are somewhat *more* fixations on the second NP, with a second peak around the 21st fixation slot which is much weaker in the match-case trials. Overall image fixation percentage is lower for the mismatch trials than for the match trials for the 12th and all subsequent fixation slots.

An interesting aspect of the match case trials is that, from the 16th to the 19th fixation slot, there is seems to be some preference to inspect the agent rather than the patient picture. In the 16th and 17th slot, there is also a raised amount of fixations on the N-CHAR region. For the mismatch trials, fixations on the N-CHAR are generally (but not in the 16th / 17th slot) more than for match trials. A differentiation between agent and patient fixation percentages happens after slot 9 for the mismatch trials but only much later – but stronger – after slot 15 for the match trials. This late-occurring phenomenon is, on the other hand, absent from the data for the mismatch trials. The patient / agent differences actually get very small at the 17th slot for mismatch trials again.

If the S-CHAR is an agent of some depicted action (the agent character or the ambiguous character of the image), many fixations (from the 11th slot on) are on that character. In the mismatch case, this "agent-reinforcement" does not take place, allowing a slight general

Figure 3.6: Histograms of fixation frequency by fixation number: Totals for all (correctly answered) SVO and OVS trials, respectively. Only data for selected regions shown, see the appendix for the full data. For SVO, ambiguous character is also S-CHAR, while it is O-CHAR for OVS.

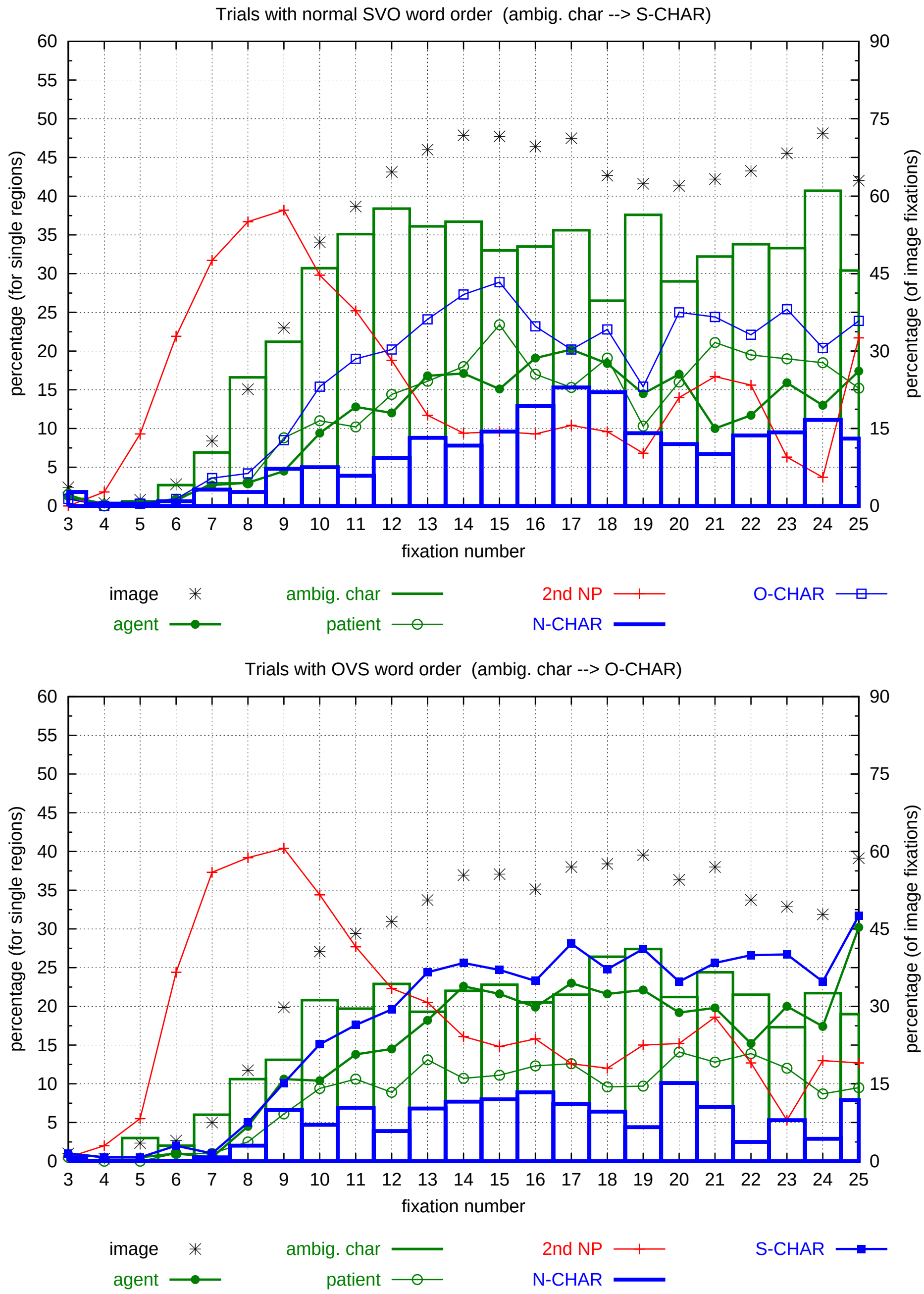
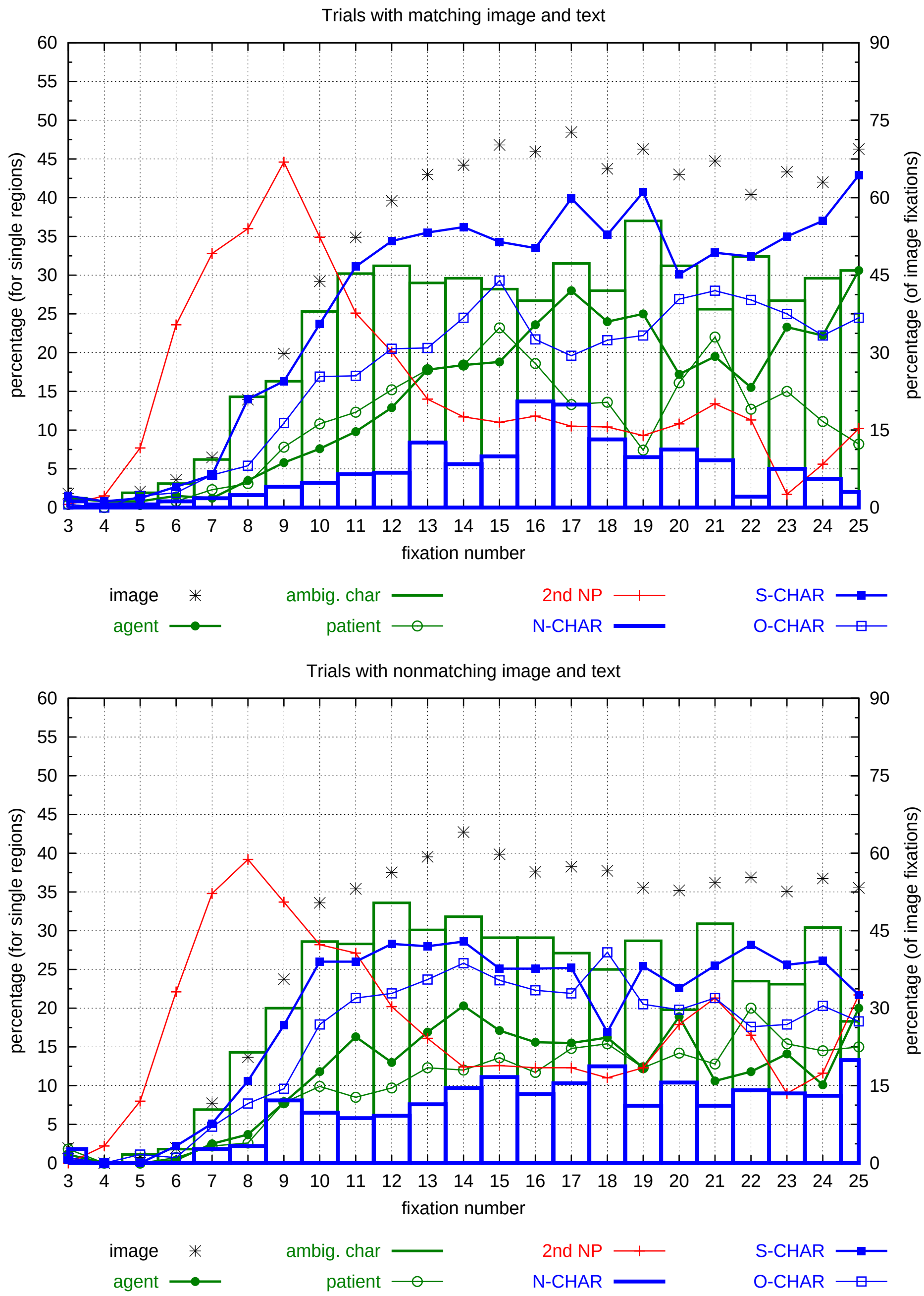


Figure 3.7: Histograms of fixation frequency by fixation number: Totals for all (correctly answered) match and mismatch trials, respectively. Only data for selected regions shown, see the appendix for the full data.



bias towards looking at the active / agent characters to become visible. In the match case, however, such a contrast only shows up much later, but stronger, as a side-effect of the strong contrast between the percentage of fixations on the S-CHAR region and on the O-CHAR region.

Having a similar big difference for the mismatch case would cause *fewer* looks at the image agent than on the image patient, as the agent / patient roles of the depicted characters contradict the roles suggested by the sentence for the mismatch case. However, preference to look at S-CHAR seems to be small in the mismatch case. Same for the preference to look at the agent picture. The differences, if they existed, might have cancelled each other out.

#### 3.2.6.4 Interactions of match / mismatch with SVO / OVS

Next, we have a look at possible *interactions* between the sentence type and the match / mismatch condition of the trials, shown in the four histograms in figure 3.8. As each histogram shows only one fixed value for match / mismatch and SVO / OVS, each histogram has a fixed mapping between agent / middle[1] / patient and S-CHAR, O-CHAR and N-CHAR. So the diagrams only show agent / ambiguous / patient character data, which can be mapped to S-CHAR, O-CHAR and N-CHAR data for each fixed trial condition: See table 3.2[2] for details about the mapping.

The SVO condition has lower fixation percentages on the second NP compared to the OVS condition. So SVO match has a narrow first peak and low percentages later, while the OVS mismatch has a streched out initial peak and the percentage of fixations on the second NP stays higher. Initial NP2 fixation percentage peaks are at slot 9 / 45% for match, slot 8 / 40% for SVO mismatch and slot 7 / 37% for OVS mismatch. The latter two could also be described as the rising pattern ending before the full peak is reached for the mismatch trials.

An interesting asymmetrical effect on the NP2 curve is that the OVS match case is not showing a second peak, while both SVO (for match: 21th / 22th slot, for mismatch: 20th to 22th slot) and OVS-mismatch (around the 21th slot) show such a peak. OVS match has some peak at the 16th slot, however. In addition, a first strong minimum occurs at the 13th (SVO match) or 14th slot, but is missing for OVS mismatch.

Both SVO and match condition come along with higher overall image fixation percentages compared to OVS and mismatch conditions respectively. This could mean that both are the more intuitive counterparts of their respective contrasting conditions. Differences start around slot 10.

Irrespective of agent / patient role, there are always very few fixations on N-CHAR: At most 17%, except for slot 17 of SVO match. Often even 10% and lower, especially for OVS match.

[1] mentioned by the first NP, which is the subject in SVO case and the object in OVS case

[2] Summary of table 3.2: For all SVO sentences, the middle depicted character is S-CHAR. For all OVS cases, it is O-CHAR. For SVO match, the patient region is at the same time the O-CHAR region. For OVS match, the agent region is at the same time the S-CHAR region. The conflict for the mismatch conditions is constructed by making the agent region O-CHAR for SVO. For OVS, the mismatch is constructed by making the patient region S-CHAR. The remaining depicted character is the N-CHAR one.

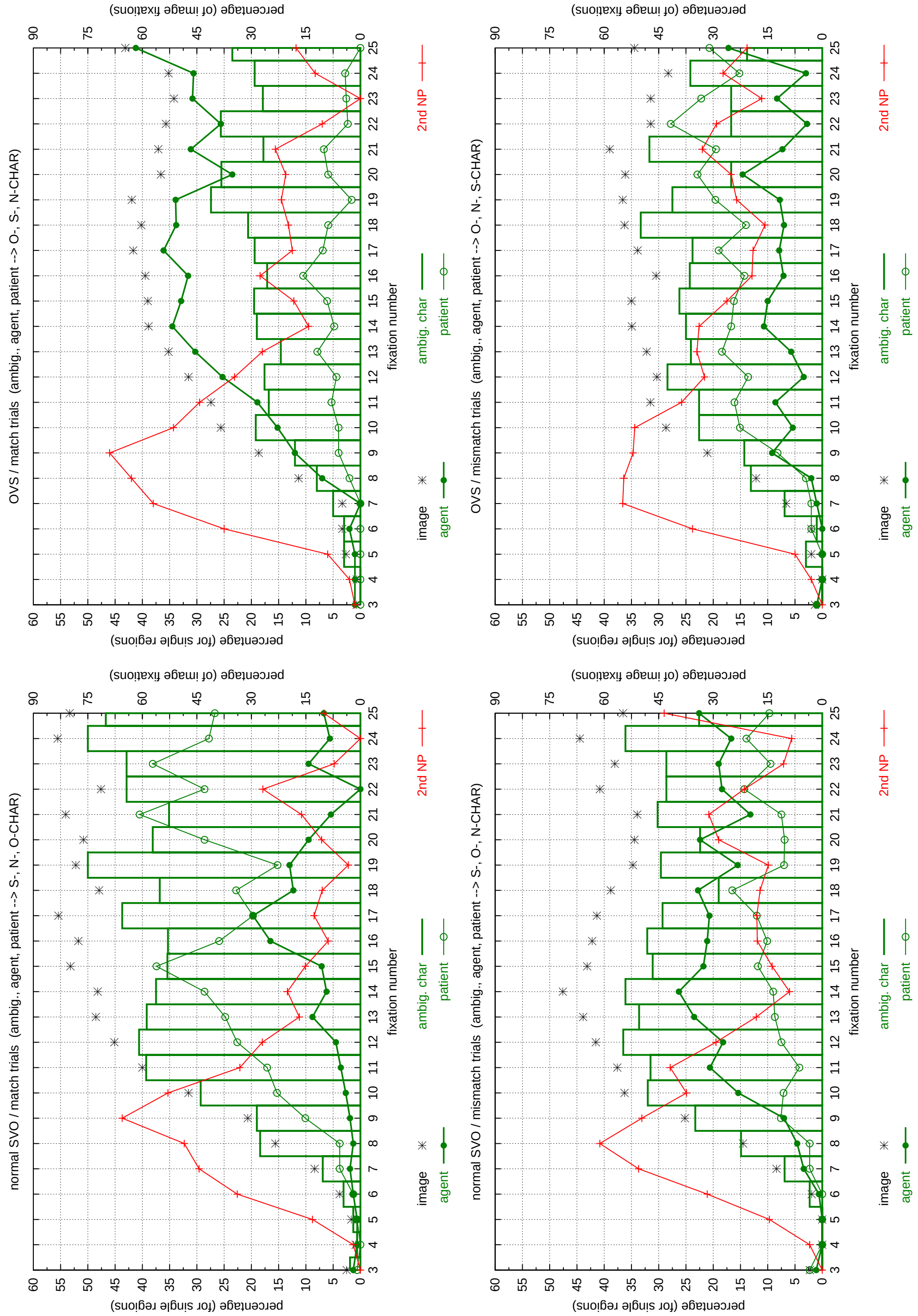



Figure 3.8: Histograms of fixation frequency by fixation number: Totals for all combinations of SVO / OVS and match / mismatch cases. Only data for selected regions shown, see the appendix for the full data.

More varied data can be seen for S-CHAR and O-CHAR regions: There are normally (not for OVS mismatch) more fixations on S-CHAR than on O-CHAR in all slots, no matter whether the regions correspond to agent, patient or ambiguous character. However, two O-CHAR peaks (slot 15, 21) for SVO match and slot 18 for SVO mismatch and slot 10 for OVS match show an opposite distribution. Notably, S-CHAR > O-CHAR only from the 11th slot on in OVS match, as opposed to from the 6th slot on for SVO.

A very interesting exception from the general pattern can be seen for OVS mismatch case: Here (with a few exceptions from the 20th slot on) O-CHAR > S-CHAR. The O-CHAR data shows a strong peak for SVO match around the 15th slot and a much weaker peak for SVO mismatch around the 14th slot. Data looks quite different for OVS. Weak O-CHAR peaks seem to exist at the 10th and 12th slot for match / mismatch there, respectively.

Describing the same data in terms of agent, ambiguous and patient character, we get: *SVO match and OVS mismatch have more agent fixations, while the other two cases have more patient fixations* – a crossover phenomenon. Bigger differences start at the 10th slot for mismatch and at the 9th / 8th slot for SVO / OVS mismatch respectively.

The ambiguous character usually received more fixations than the other two characters, but *not for OVS match*, where the agent character region received more fixations. For SVO, the ambiguous character is also S-CHAR. The ambiguous character always got more fixations than the patient, while the agent character got ranked 1st for OVS match, 2nd for SVO mismatch and 3rd for SVO match and OVS mismatch. The latter two cases are those where the agent region is also the N-CHAR region.

If we exclude the N-CHAR from the comparison, we get agent > ambiguous > patient with the exception being SVO mismatch, where we have ambiguous > agent. Reasoning in terms of NP1-CHAR (always same region as the ambiguous character) and NP2-CHAR and N-CHAR, we get NP1-CHAR > NP2-CHAR > N-CHAR. Yet again there is an exception: For OVS match, NP2-CHAR > NP1-CHAR. We will later (in the discussion) present a set of rules which describes the observed data for all four trial conditions.

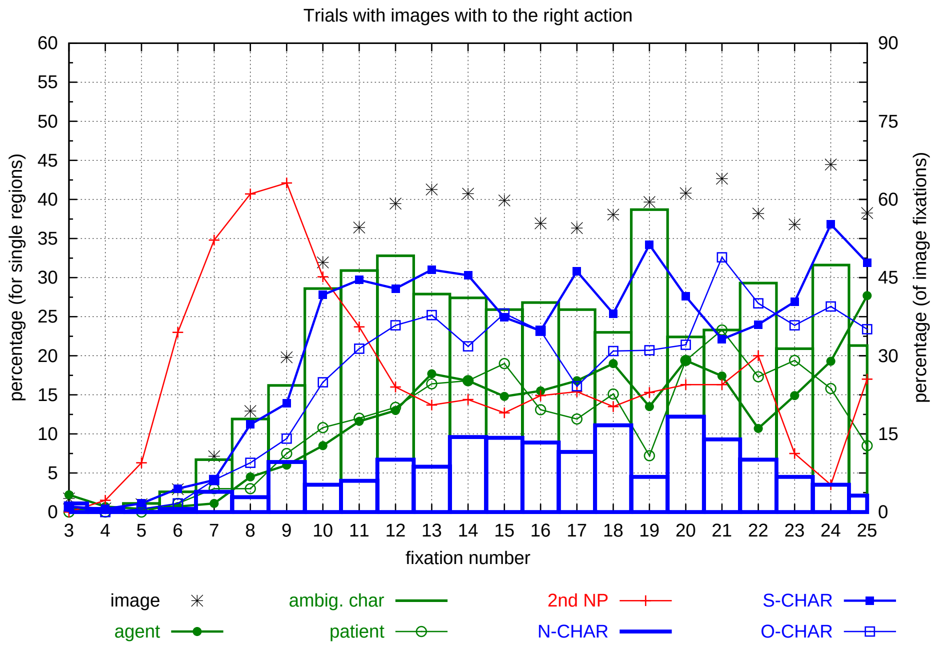
#### 3.2.6.5 Left versus right image action direction

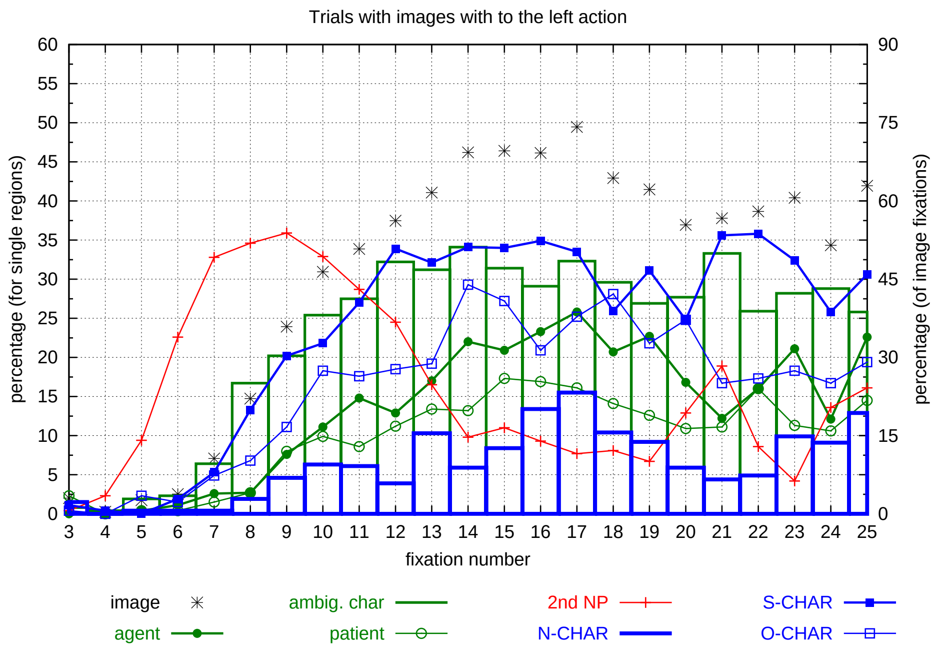
Finally, we look at the differences related to the direction (to the left or to the right) of the depicted actions. That contrast is not expected to have strong effects, so it is not analyzed in context of the other (sentence type and match / mismatch) contrasts. Those who are interested in the fine details can find some additional histograms about interactions between direction and other factors in appendix B.

As the left / right contrast is orthogonal to the others in terms of image item labelling, figure 3.9 has to show agent, ambiguous character (mentioned by the first NP) patient, S-CHAR, O-CHAR and N-CHAR data all separately. In addition, fixation percentages for the second NP and for image regions in general are shown.

Trials which have images with action to the right have a higher peak for the second NP text region, with quicker but less deep (only to 15% rather than to 10%) falloff. If the image action direction is to the left, there are more fixations on image regions in slots 14 to 18. There are

Figure 3.9: Histograms of fixation frequency by fixation number: Totals for all (correctly answered) trials with to-the-left and to-the-right image action direction, respectively. Only data for selected regions shown, see the appendix for the full data.



also more agent than patient fixations in slots 11 and later for image action direction to the left. Both (more image fixations, difference agent / patient) phenomena possibly correspond to a "circular" eye movement pattern which we saw in the data, using DPIFilter's visualization component: Start in the middle (forced by experiment design), read the text in the upper region of the screen from the left to the right, and inspect the image at the bottom of the screen from the right to the left. This pattern seems to fit best if the agent is on the right / if the image action direction is to the left. When the circular pattern fits the image action direction, slightly more fixations to the N-CHAR can be observed as well.

For both to-the-left and to-the-right image action direction, fixation proportions on the S-CHAR are higher than those on the O-CHAR region, but some slight irregular interaction with the image action direction seems to be present here, too.

### 3.2.7 Inferential analysis

#### 3.2.7.1 List of analyzed parameters

Four sentence regions (plus the background) of the stimulus display are analyzed. See sections 3.1.4 and 3.1.4.3 for details about regions and analysis. Three image regions (corresponding to the depicted characters) are labelled and analyzed in three different ways each: In terms of agent / ambiguous / patient character, in terms of a sentence based point of view as S-CHAR / O-CHAR / N-CHAR, and – to check for direction / balancing related asymmetries – by position (left / middle / right). Middle and ambiguous character are always the same. Together with the three extra region *groups* "image" (any depicted character) and "sentence" and "anywhere at all" (pool of all fixations of a trial), we analyzed data for 16 distinct regions.

For each of the 16 regions the three basic measures of fixation count, total fixation duration and first fixation duration were calculated, using a custom Perl script[1]. We also calculated the measures for "first inspection" (two measures: duration and number of fixations) for the 8 character region labels. For each of the four text regions, we calculated measures for "regression path" (fixation count and duration of fixations) and "first pass" (same values). The "first inspection", "regression path" and "first pass" are clusters of fixations, selected by rules explained in section 3.1.4.3.

For all measures (see table 3.3), the per-condition sample means and standard deviations were calculated (again using a custom Perl script). The conditions are: SVO, OVS, match, mismatch, plus the balancing cases to-left and to-right.

In preparation for some additional post-hoc analysis, we also evaluated all three sets of possible four-way splits (SVO match, SVO mismatch, OVS match, OVS mismatch as main set, plus two sets which combine direction balancing with either SVO / OVS or match / mismatch).

Our Perl script processed the data for all correctly answered trials. There were many incorrectly answered trials (in particular for OVS conditions), so we also did the same calculations on the incorrectly answered trials as well, creating a separate data set.

[1]All Perl scripts and gnuplot scripts used for this thesis were written by the author and are available on request.

#### 3.2.7.2 The effect search method, the table format, some examples

Based on this data, a systematic search for significant differences along some contrasts of interest was performed: Two-tailed t-tests are used to check along which of the contrast axes (SVO / OVS, match / mismatch, left / right image action direction) effects exist. We expect no particular effects along the left / right balancing contrast, of course (except for the obvious left / right character regions, for which the image action direction determines the agent / patient role assignment).

In each of the following tables, we present results for one or two of the contrast axes, starting with a combined table for SVO / OVS and match / mismatch. This table therefore answers the two questions: Is there an effect of SVO / OVS? Is there an effect of match / mismatch? Answers are shown for each measure where at least one of both questions is found to have the answer *yes*. In other words, if no effects along either of both axes could be detected for a measure, then the data for that measure is not shown in the table at all. All measures (see table 3.3) are processed for generating the tables, so if a particular measure is not listed (for a specific region, or even not listed at all), then we did not find any significant effects of the trial conditions on that measure. No lines have been manually removed.

The tables also show results for a post-hoc Exploratory Data Analysis. If for example the main axes of SVO / OVS and match / mismatch are shown in a table, the EDA answers the following questions about sub-effects and ordered means: Is there an effect of match / mismatch if we look only at the SVO trials? Or only the OVS trials? Conversely, is there an effect of SVO / OVS if we look only at the match trials? Or only the mismatch trials? If we sort the conditions by their average measures (e.g. "which condition has the fewest text fixations?"), what kind of partial ordering do we get? All this is done by a set of two-tailed t-tests for all six pairs of the four conditions selected by the contrast axes of the table. In the example, the conditions are SVO match, SVO mismatch, OVS match and OVS mismatch.

Because the EDA tests are post-hoc, significance levels have to be treated with care: The primary ordered-means analysis uses results from up to six out of the six possible comparison pairs between our four conditions to select symbols which describe the partial ordering / effects (see table 3.6). The four "Is there an effect of ... if we only look at...?" questions use four of the same six comparisons.

The risk of getting false positives ("spurious effects") increases linearly with the number of pairwise t-tests performed. So given that our EDA uses some subset of six t-tests, a conservative correction would be to require significances to have $p < (0.05/6)$ to avoid seeing spurious effects. We decided to simply drop EDA t-test results where significances did not reach $p < 0.01$ or better. This is slightly over-conservative for the four sub-effect tests[1] and slightly under-conservative for the "what kind of partial ordering...?" symbol selection, but we still have a $p < 0.06$ worst case there. Some aspects of EDA are explained in the introduction [@NYU]. Tukey has written a book titled EDA, [Tuk77].

[1] We selected the fixed subset of four of the six pairwise t-tests in advance, so even though the EDA is post-hoc, it would arguably already be completely conservative / safe to correct the threshold by factor 4 instead of factor 6 as far as the four sub-effect table columns are concerned.

**An example of the table format:** Table 3.7 starts with data about button time[1] and so the first table row shows that there are highly significant effects of SVO / OVS sentence type and that the match / mismatch distinction only has a significant effect on SVO sentences. It also shows that the effects of SVO / OVS sentence type are not significant if only mismatch trials are analyzed.

The rightmost column presents the ordered (by value) listing of mean values for each of the four compared conditions of the search for each of the result table lines. In the button time example, the mean time is 4605 milliseconds for the SVO / match case, 5164 milliseconds for SVO / mismatch, 5928 for OVS / mismatch and 6324 for OVS / match. Of course mean values alone do not reveal whether there is a really significant difference between them. The three symbols between the four mean values indicate whether and where the EDA found significant contrasts: See table 3.6 for an explanation of the symbols.

To keep the tables short, rows are only shown when they show at least one main effect or one sub-effect. If there is some effect *except* $\prec$ or $\circ$ presented in the rightmost column, the row is shown as well. The rightmost column trigger is limited to real-$\bullet$ and $\triangleleft$ because the other symbols imply that at least one sub-effect exists, and we generally want to control row inclusion via sub-effects, not via the symbols in the rightmost column. This may seem redundant but makes sense for the image action direction related tables, where only one trial factor axis (the image action direction) is used for triggering. Main effects along the other axis are already discussed in other tables and therefore not shown again in the image action direction related tables. However, we do show *asymmetrical* (i.e. influenced by the image action direction) sub-effects of the direction in such tables. Note that such asymmetries are *no significant effects* in themselves – they are only shown as supplementary informal description of our data.

For the button time, the inferential EDA reveals the following, as shown in table 3.7: Answers to SVO match stimuli were significantly faster than those to SVO mismatch and to OVS stimuli. Answers to OVS mismatch stimuli only took insignificantly longer than answers to

Table 3.6: Explanation of the symbols used in the sorted averages column of the effect tables

| | |
|---|---|
| $\prec$ | All values left of the $\prec$ are significantly smaller than all values right of it (with $p < 0.01$). |
| $\triangleleft$ | The value directly left of the $\triangleleft$ is significantly smaller the one right of it with $p < 0.01$, but $\prec$ does not hold at this place. |
| $\bullet$ | Non-local significant difference: When shown in left / middle / right slot respectively, difference is between 1st and 3rd / 1st and 4th / 2nd and 4th value, respectively. Only show where $\triangleleft$ and $\prec$ do not already hold for the slot. |
| $\circ$ | Substitutes $\bullet$ if $\prec$ holds at an *adjacent* slot. This is to indicate that the non-local difference is implied by the adjacent $\prec$. |
| $\wr$ | Neither $\prec$ nor $\triangleleft$ nor $\bullet$ nor $\circ$ holds at this place. |

[1]Button time is the response time between stimulus onset and button press.

SVO mismatch stimuli (actually $p < 0.05$ but not $p < 0.01$, so the effect is not reliable in context of the post-hoc multiple pairwise t-test EDA and is filtered from the table). There is a main effect of SVO / OVS on answer time, but no main effect of congruence (match / mismatch) on answer time. Nevertheless, we saw that the EDA results show some interactions between SVO / OVS and congruence. There is a significant sub-effect of congruence among SVO trials.

You will have noticed that the table does not list the average values for button time for SVO or for OVS or for match or for mismatch cases separately. You have to refer to the appendix to find that they are 4898 milliseconds for SVO and 6125 milliseconds for OVS. As there is no significant overall difference between match and mismatch trials for the button time, not even the appendix lists the average button times for match and mismatch, respectively. The table in this chapter does list the average values for the four basic conditions (here: SVO match, SVO mismatch, OVS match, OVS mismatch), though: Those are shown as part of the "sorted averages" EDA column of the table. As sample sizes vary, you must refer to the sample size information in the appendix if you want to calculate combined *weighted* averages. Example: Combine averages for SVO match and OVS match to get the overall match average.

The appendix also lists the t-values for each of the performed tests whenever a t-test found a significance better than the threshold. For example SVO and OVS button times differ with $p < 0.001$ at $t = 5.886$ (if you only look at the match cases: $p < 0.001$ at $t = 6.005$). Looking only at the SVO cases, times for match and mismatch differ with $p < 0.01$ at $t = 2.686$, while the overall match / mismatch difference does not reach significance of $p < 0.05$ and is not presented in detail in the appendix.

#### 3.2.7.3 Effects of sentence type and congruence

After the in-depth introduction to the table format used in the following (sub-)sections, we can now focus on the contrasts shown in the first and most important effect table, the one for the two main contrasts SVO / OVS and match / mismatch, table 3.7.

We did not explicitly analyze the *per fixation* durations, but we do present some manually calculated values here: The average durations range from 178 to 186 msec for background, from 215 to 221 msec for text and from 254 to 265 msec for image regions, depending on condition. The global average of per fixation duration is 230 to 234 msec depending on trial condition. Some interesting condition-related differences: Sentence SVO 215 msec, OVS 219 to 221 msec. Agent: SVO match 238, OVS mismatch 230, SVO mismatch 261, OVS match 283 msec. The interesting detail here is that SVO match / OVS mismatch differ in the opposite direction in per-fixation-duration compared to their respective fixation count and (summed) fixation duration data. Patient: OVS match 241, SVO mismatch 245, SVO match 259, OVS mismatch 285 msec. Middle ambiguous character[1]: OVS match 259, mismatch 263, SVO match 267, mismatch 272 msec. S-CHAR OVS 283 to 284 msec, SVO 273 msec – while overall count and duration show a different pattern, as we will see later. O-CHAR SVO 258 to 261 msec, OVS 263 to 267 msec (267 for match), N-CHAR 230 to 245 msec (OVS mismatch 230, others at least 238 msec). Values are – for all conditions – N-CHAR < O-

[1]Very first fixation of each trial excluded, as usual

CHAR < S-CHAR. First NP: SVO match 199, mismatch 205, OVS match 205, mismatch 209 msec. Verb: SVO mismatch 217, match 224 msec (difference is in other direction than fixation count / duration), OVS mismatch 230, match 228 msec. Adverb: 209 to 217 msec (209 for OVS match, at least 215 for the others). Second NP: SVO match 225 msec, SVO mismatch 221 msec, OVS match 227 msec, OVS mismatch 234 msec. We did not check variances of per fixation durations, so we make no claims about the significance of any of the differences between the conditions.

As mentioned in the example for the table format, participants answered significantly faster for SVO trials than for OVS trials. However, the difference is not significant (threshold for post-hoc test is $p < 0.01$) among mismatch trials. The difference between match and mismatch trials themselves is not significant in the overall case, only for SVO trials, where the SVO match decision is faster than the SVO mismatch decision. For OVS trials, the difference in answer times for match and mismatch is too small to be significant. Confirming OVS match takes *insignificantly longer* than confirming OVS mismatch. As a more or less trial condition independent percentage of the overall trial time is part of fixations in our data, the anywhere fixation time effects show exactly the same pattern as the button time. Similarily, the anywhere fixation count, on the average depending in a quite linear fashion on the overall fixation time, again shows the same effects as the button time.

The background fixations depend on the sentence type, too. This could be a side effect of the very big overall trial duration effect of sentence type: If some percentage of the fixations is on the background, then the background area data shows similar effects than the overall trial duration. A similar effect shows up for the "image" region group, but here the contrast focuses on the match cases: Confirming a SVO match needs less inspection time and fixations than confirming an OVS match.

The "sentence" region (group) shows a different picture: Here, answering YES involves fewer fixations and less fixation time than answering NO, but only significantly so for the SVO case. The OVS case generally involves many text inspections which do not get significantly more for a mismatch.

Now we look at effects in the individual text regions: For the first NP and for the verb, there are fewer fixations for SVO than for OVS (main effect originating from a match sub-effect here). For the adverbial region, a similar main effect of SVO / OVS is present and (other than for NP1 / verb) it is not limited to match cases. This could be an artifact caused by some fixations on NP2 being accidentally counted as fixations on the adverbial region. The NP2 region itself has strong effects of SVO / OVS sentence type (SVO "faster" than OVS). The NP2 region is also the only text region for which at least sub-effects of match / mismatch reach significance levels of $p < 0.01$ or better – here for the SVO cases, for the regression path. None of the other text regions have regression path effects. All text regions have (in context of post-hoc analysis thresholds) insignificant sub-effects of match / mismatch: $p < 0.05$ but not $p < 0.01$ among SVO trials.

For the second NP region, fewest and shortest regression path fixations happen in the SVO match case, more in the SVO mismatch case, and even more in the OVS cases, all being significant differences found by post-hoc exploratory data analysis.

None of the text regions exhibits first pass effects. So all effects are caused by fixations *after* the first pass and are effects of re-reading. An expected result, as the participants cannot know whether the sentence is SVO or OVS before they see the determiner of the second NP. The image regions do show first inspection effects, though: As we can see in the histograms, most image inspections took place after reading the text, so it is not too surprising to see first inspection effects for the image regions.

Labelling image regions by their depicted role gives results in a crossover pattern for the agent character region: For match cases, SVO is "faster" (less fixations / shorter total fixation duration / shorter first fixation time / shorter first inspection time / fewer fixations in the first inspection) than OVS. For mismatch cases, OVS is "faster" than SVO for the agent region. There is *no* significant difference in agent region fixation data between match and mismatch cases in general, but it *is* the case that YES answers require fewer agent fixations than NO answers for SVO trials. For OVS trials, YES answers take *more* agent fixations than NO answers.

The agent region effects can be summarized in a simple rule: *The agent character gets the fewest fixations when it is N-CHAR (SVO match case and OVS mismatch case), more fixations when it is O-CHAR (SVO mismatch case) and the most and longest fixations when it matches the S-CHAR region.* All agent region effects are highly significant ($p < 0.001$). The average fixation times for the three sentence roles of the agent image are less than 0.3 seconds, about 0.6 seconds and about 1.2 seconds respectively, having a factor of two at each step.

The patient region data shows a similar pattern: Yet again, the two cases where the patient character is not mentioned in the text correspond to the shortest and fewest fixations. However, the OVS mismatch (patient is S-CHAR) and the SVO match (patient is O-CHAR) cases now exhibit an almost *equal* amount of fixations to the patient image. So the patient picture is less receptible to effects of the corresponding sentence role than the agent picture. In addition, the patient picture shows no overall SVO / OVS effect, simply because the SVO / OVS effects for match and mismatch still have opposite directions but now have similar strength. Both sub-effects cancel each other out, so there is no main effect. The match / mismatch sub-effect for SVO and OVS effects cancel each other out for *both* the agent and the patient image.

The agent / patient distinction turned out to be interesting, but we also think that it does not capture another important point of view well: What seemed to be an effect of sentence type or match / mismatch seems to be mostly a constant difference linked to the sentence role, a "fixation popularity ranking" with unmentioned weakest, mentioned by sentence object stronger and mentioned by sentence subject strongest. Therefore we now change the point of view to the S-CHAR / O-CHAR / N-CHAR labelling of the image areas.

We now see that, apart from the constant factors related to the sentence role of a picture, there is in fact *some* effect of SVO / OVS and match / mismatch for the S-CHAR and O-CHAR regions: In the overall scope, there are shorter and less looks at the O-CHAR for the SVO case. The post-hoc analysis also detects shorter and less fixations at the S-CHAR for the OVS mismatch case. A main effect for S-CHAR only exists for match / mismatch but not for sentence type. The match / mismatch effect originates almost entirely from OVS trials as the post-hoc analysis confirms. For SVO trials, match is insignificantly "faster" than mismatch.

Table 3.7: Significance of SVO / OVS and match (√) / mismatch (⋆) effects and interactions of both

| **Region / measure** | Is SVO ≉ OVS? over-all | Is SVO ≉ OVS? only for √ | Is SVO ≉ OVS? only for ⋆ | match ≉ mismatch? over-all | match ≉ mismatch? only for SVO | match ≉ mismatch? only for OVS | **Averages**, sorted (S/Y is SVO/match, O/N is OVS/mismatch) |
|---|---|---|---|---|---|---|---|
| **button** | | | | | | | |
| time | 0.001 | 0.001 | — | — | 0.01 | — | S/Y 4605 ≺ S/N 5164 ○ O/N 5928 • O/Y 6324 |
| **anywhere** | | | | | | | |
| time | 0.001 | 0.001 | — | — | 0.01 | — | S/Y 3921 ≺ S/N 4447 ○ O/N 5104 • O/Y 5439 |
| # of fix. | 0.001 | 0.001 | — | — | 0.01 | — | S/Y 16.99 ≺ S/N 19.34 ○ O/N 21.89 • O/Y 23.22 |
| **background** | | | | | | | |
| time | 0.01 | — | — | — | — | — | S/Y 157.4 ≀ S/N 180.9 • O/Y 241.8 ≀ O/N 250.6 |
| # of fix. | 0.001 | 0.01 | — | — | — | — | S/Y 0.843 • S/N 1.017 • O/Y 1.300 ≀ O/N 1.376 |
| **sentence** | | | | | | | |
| time | 0.001 | 0.001 | 0.01 | — | 0.01 | — | S/Y 1857 ≺ S/N 2181 ≺ O/Y 2695 ○ O/N 2726 |
| # of fix. | 0.001 | 0.001 | 0.01 | — | 0.01 | — | S/Y 8.635 ≺ S/N 10.15 ≺ O/N 12.20 ○ O/Y 12.47 |
| **first NP** | | | | | | | |
| time | 0.001 | 0.001 | — | — | — | — | S/Y 504.9 • S/N 599.2 • O/N 696.2 ≀ O/Y 714.2 |
| # of fix. | 0.001 | 0.001 | — | — | — | — | S/Y 2.541 • S/N 2.920 • O/N 3.337 ≀ O/Y 3.480 |
| **verb** | | | | | | | |
| time | 0.01 | 0.001 | — | — | — | — | S/Y 400.1 • S/N 454.5 • O/N 507.4 ≀ O/Y 550.3 |
| # of fix. | 0.01 | 0.001 | — | — | — | — | S/Y 1.786 • S/N 2.091 • O/N 2.208 ≀ O/Y 2.410 |
| **adverb** | | | | | | | |
| time | 0.001 | 0.001 | 0.01 | — | — | — | S/Y 378.1 • S/N 457.9 • O/Y 557.8 • O/N 608.9 |
| # of fix. | 0.001 | 0.001 | 0.01 | — | — | — | S/Y 1.761 • S/N 2.114 • O/Y 2.670 • O/N 2.812 |
| **second NP** | | | | | | | |
| time | 0.001 | 0.001 | 0.001 | — | — | — | S/Y 574.3 ○ S/N 669.2 ≺ O/Y 873.0 ○ O/N 913.6 |
| # of fix. | 0.001 | 0.001 | 0.01 | — | — | — | S/Y 2.547 ○ S/N 3.023 ≺ O/N 3.842 ○ O/Y 3.910 |
| r-path time | 0.001 | 0.001 | 0.001 | — | 0.01 | — | S/Y 819.4 ≺ S/N 1104 ≺ O/Y 1651 ○ O/N 1706 |
| r-path # | 0.001 | 0.001 | 0.001 | — | 0.01 | — | S/Y 3.654 ≺ S/N 5.040 ≺ O/Y 7.370 ○ O/N 7.376 |
| **image** | | | | | | | |
| time | 0.05 | 0.001 | — | — | — | — | S/Y 1906 ≀ S/N 2085 • O/N 2127 ≀ O/Y 2502 |
| # of fix. | 0.05 | 0.01 | — | — | — | — | S/Y 7.516 ≀ S/N 8.171 • O/N 8.317 ≀ O/Y 9.450 |
| **agent** | | | | | | | |
| time | 0.001 | 0.001 | 0.001 | — | 0.001 | 0.001 | S/Y 182.6 ○ O/N 264.4 ≺ S/N 622.9 ≺ O/Y 1224 |
| # of fix. | 0.001 | 0.001 | 0.001 | — | 0.001 | 0.001 | S/Y 0.767 ≺ O/N 1.149 ≺ S/N 2.389 ≺ O/Y 4.320 |

Continued on next page

Table 3.7 – continued from previous page

| Region / measure | SVO $\not\approx$ OVS both | SVO $\not\approx$ OVS $\surd$ | SVO $\not\approx$ OVS $\star$ | match $\not\approx$ mismatch both | match $\not\approx$ mismatch SVO | match $\not\approx$ mismatch OVS | Order of values |
|---|---|---|---|---|---|---|---|
| 1st fix. time | 0.05 | 0.001 | 0.001 | — | 0.001 | 0.001 | S/Y 108.7 ∘ O/N 131.4 ≺ S/N 203.4 ≺ O/Y 254.6 |
| 1st pass time | 0.001 | 0.001 | 0.001 | — | 0.001 | 0.001 | S/Y 146.1 ∘ O/N 204.1 ≺ S/N 369.4 ≺ O/Y 667.3 |
| 1st pass # | 0.001 | 0.001 | 0.001 | — | 0.001 | 0.001 | S/Y 0.629 ∘ O/N 0.881 ≺ S/N 1.429 ≺ O/Y 2.360 |
| **patient** | | | | | | | |
| time | — | 0.001 | 0.001 | — | 0.001 | 0.001 | O/Y 205.2 ∘ S/N 287.4 ≺ S/Y 575.8 ∘ O/N 694.0 |
| # of fix. | — | 0.001 | 0.001 | — | 0.001 | 0.001 | O/Y 0.850 ∘ S/N 1.171 ≺ S/Y 2.226 ∘ O/N 2.436 |
| 1st fix. time | — | 0.001 | 0.001 | — | 0.001 | 0.001 | O/Y 119.4 ∘ S/N 158.4 ≺ S/Y 227.6 ∘ O/N 253.9 |
| 1st pass time | — | 0.001 | 0.001 | — | 0.001 | 0.001 | O/Y 150.4 ∘ S/N 189.9 ≺ S/Y 324.3 ∘ O/N 372.3 |
| 1st pass # | — | 0.001 | 0.001 | — | 0.001 | 0.001 | O/Y 0.650 ∘ S/N 0.777 ≺ S/Y 1.277 ∘ O/N 1.307 |
| **S-CHAR** | | | | | | | |
| time | — | — | 0.01 | 0.05 | — | 0.001 | O/N 694.0 ≺ S/Y 977.6 ∘ S/N 1008 ≀ O/Y 1224 |
| # of fix. | — | — | 0.001 | 0.01 | — | 0.001 | O/N 2.436 ≺ S/Y 3.585 ∘ S/N 3.691 ≀ O/Y 4.320 |
| 1st pass time | — | 0.01 | 0.01 | 0.05 | — | 0.001 | O/N 372.3 ≺ S/Y 486.4 ∘ S/N 525.0 • O/Y 667.3 |
| 1st pass # | — | 0.01 | 0.001 | 0.01 | — | 0.001 | O/N 1.307 ≺ S/Y 1.824 ∘ S/N 1.931 • O/Y 2.360 |
| **O-CHAR** | | | | | | | |
| time | 0.001 | 0.001 | 0.001 | — | — | — | S/Y 575.8 ∘ S/N 622.9 ≺ O/Y 901.2 ∘ O/N 993.9 |
| # of fix. | 0.001 | 0.001 | 0.001 | — | — | — | S/Y 2.226 ∘ S/N 2.389 ≺ O/Y 3.370 ∘ O/N 3.782 |
| 1st pass time | — | — | — | 0.05 | — | — | S/Y 324.3 ≀ O/Y 329.0 ≀ S/N 369.4 ≀ O/N 411.8 |
| 1st pass # | — | — | — | 0.01 | — | — | O/Y 1.270 ≀ S/Y 1.277 ≀ S/N 1.429 • O/N 1.564 |
| **N-CHAR** | | | | | | | |
| time | — | — | — | 0.01 | 0.01 | — | S/Y 182.6 ≀ O/Y 205.2 • O/N 264.4 ≀ S/N 287.4 |
| # of fix. | — | — | — | 0.01 | 0.01 | — | S/Y 0.767 • O/Y 0.850 • O/N 1.149 ≀ S/N 1.171 |
| 1st fix. time | — | — | — | 0.01 | 0.01 | — | S/Y 108.7 ≀ O/Y 119.4 • O/N 131.4 ≀ S/N 158.4 |
| 1st pass time | — | — | — | 0.01 | — | — | S/Y 146.1 ≀ O/Y 150.4 ≀ S/N 189.9 ≀ O/N 204.1 |
| 1st pass # | — | — | — | 0.01 | — | — | S/Y 0.629 ≀ O/Y 0.650 ≀ S/N 0.777 ≀ O/N 0.881 |
| **left pic.** | | | | | | | |
| time | 0.001 | 0.001 | — | — | — | — | S/Y 371.4 ≀ S/N 457.9 • O/N 524.3 • O/Y 736.2 |
| # of fix. | 0.01 | 0.001 | — | — | — | — | S/Y 1.403 ≀ S/N 1.777 • O/N 1.861 • O/Y 2.510 |
| 1st fix. time | 0.05 | 0.01 | — | — | — | — | S/Y 160.6 ≀ S/N 183.6 • O/N 202.8 ≀ O/Y 220.0 |
| 1st pass time | 0.001 | 0.001 | — | — | — | — | S/Y 217.9 ≀ S/N 284.5 • O/N 295.9 • O/Y 448.2 |
| 1st pass # | 0.01 | 0.001 | — | — | — | — | S/Y 0.874 ≀ O/N 1.069 • S/N 1.103 ◁ O/Y 1.540 |
| **middle pic.** | | | | | | | |
| 1st pass time | 0.001 | 0.001 | — | — | — | — | O/Y 329.0 • O/N 411.8 • S/Y 486.4 ≀ S/N 525.0 |
| 1st pass # | 0.001 | 0.001 | — | 0.05 | — | — | O/Y 1.270 • O/N 1.564 • S/Y 1.824 ≀ S/N 1.931 |

Continued on next page

Table 3.7 – continued from previous page

| Region / measure | SVO ≉ OVS both | SVO ≉ OVS ✓ | SVO ≉ OVS ⋆ | match ≉ mismatch both | match ≉ mismatch SVO | match ≉ mismatch OVS | Order of values |
|---|---|---|---|---|---|---|---|
| **right pic.** | | | | | | | |
| time | 0.01 | 0.001 | — | — | — | — | S/Y 386.9 ≀ O/N 434.1 • S/N 452.3 ≀ O/Y 692.8 |
| # of fix. | 0.05 | 0.01 | — | — | — | — | S/Y 1.591 ≀ O/N 1.723 • S/N 1.783 ◁ O/Y 2.660 |
| 1st pass time | 0.05 | — | — | — | — | — | S/Y 252.5 ≀ S/N 274.9 ≀ O/N 280.5 ≀ O/Y 369.5 |
| 1st pass # | 0.05 | 0.01 | — | — | — | — | S/Y 1.031 ≀ S/N 1.103 • O/N 1.119 ≀ O/Y 1.470 |

fix.: fixation(s), #: number (of fixations), ins.: inspection, r-path: regression path

Underlined: effect in "nonstandard" direction (OVS ≺ SVO or mismatch ≺ match)

The OVS match / mismatch effect has the opposite direction and is highly significant. An interesting detail is that the *per fixation* durations (not shown in the table) are consistently shorter for SVO (average 273 msec) than for OVS (average 283 to 284 msec). So there might actually be a main effect of SVO / OVS in *that* measure, but our custom Perl script does not yet analyze the per fixation durations[1] and fails to detect an effect.

For the N-CHAR region, there are more and longer fixations when the task is to confirm a mismatch compared to the task of confirming a match. However, the main effect is only backed by the SVO case, not by OVS case, as the analysis of sub-effects shows.

Using the left / middle / right distinction: There are main effects of sentence type, but sub-effects are only significant for the match case. The left / right characters get fewer fixations / shorter inspection for SVO trials than for OVS trials. Effects on the middle ambiguous character regions are, interestingly, limited to first inspection data (no overall fixation count or duration effects). First inspection is longer for SVO than for OVS here.

An explanation for the left / ambiguous / right character region effects can be taken from the fact that the left / right regions have a 50% chance to coincide with the S-CHAR for OVS and with the O-CHAR for SVO. The ambiguous character is S-CHAR for SVO and O-CHAR for OVS. We saw earlier that S-CHAR receives more fixations / longer inspections than the O-CHAR region in match cases.

#### 3.2.7.4 Effects of image action direction

While we did not originally intend to detect effects of the balancing factor of image action direction, the orthogonal design of our stimuli allows us to check for such effects as well. Main purpose of having a mix of both to the left and to the right image action direction is balancing, to allow possible left / right preferences to be cancelled out by presenting a mix of stimuli with equal numbers of instances of both image action directions.

[1] We suggest to take both in-trial and between-trials variance into account when analyzing per fixation duration data. The measures which we already do analyze do not involve in-trial variance.

Table 3.8 shows the effects of image action direction and their interaction with the SVO / OVS contrast: There are, as expected, only few main effects. As the image action direction determines the assignment of agent and patient to left and right character, effects on left / right character regions are no surprise and are not discussed in detail here.

Table 3.8: Significance of to-right / to-left effects and left / right asymmetries in SVO / OVS effects.

| **Region** / **measure** | Is $\Rightarrow \not\approx \Leftarrow$? over-all | only for SVO | only for OVS | Is SVO $\not\approx$ OVS? significances for $\Rightarrow$ | for $\Leftarrow$ | **Averages**, sorted (R/S is $\Rightarrow$/SVO, L/O is $\Leftarrow$/OVS) |
|---|---|---|---|---|---|---|
| **anywhere** | | | | | | |
| # of fix. | — | — | — | 0.001 | 0.01 | R/S 17.83 $\circ$ L/S 18.62 $\prec$ L/O 22.53 $\circ$ R/O 22.58 |
| **background** | | | | | | |
| time | — | — | — | — | 0.01 | L/S 152.4 $\bullet$ R/S 186.9 $\bullet$ R/O 233.9 $\wr$ L/O 259.0 |
| # of fix. | — | — | — | — | 0.001 | L/S 0.880 $\wr$ R/S 0.988 $\bullet$ R/O 1.245 $\bullet$ L/O 1.434 |
| **sentence** | | | | | | |
| # of fix. | — | — | — | 0.001 | 0.01 | R/S 9.327 $\circ$ L/S 9.530 $\prec$ L/O 11.68 $\circ$ R/O 12.97 |
| **first NP** | | | | | | |
| time | — | — | — | 0.01 | — | L/S 543.2 $\wr$ R/S 565.3 $\bullet$ L/O 662.5 $\bullet$ R/O 746.6 |
| # of fix. | — | — | — | 0.01 | — | R/S 2.732 $\wr$ L/S 2.747 $\bullet$ L/O 3.293 $\bullet$ R/O 3.520 |
| 1st fix. time | 0.05 | — | — | — | — | L/S 175.0 $\wr$ L/O 180.9 $\wr$ R/O 190.6 $\wr$ R/S 194.4 |
| **verb** | | | | | | |
| time | — | — | — | 0.01 | — | L/S 416.0 $\wr$ R/S 441.1 $\bullet$ L/O 491.4 $\bullet$ R/O 565.0 |
| **adverb** | | | | | | |
| time | — | — | — | 0.001 | — | R/S 407.6 $\bullet$ L/S 432.2 $\bullet$ L/O 523.8 $\bullet$ R/O 641.3 |
| # of fix. | — | — | — | 0.001 | — | R/S 1.857 $\bullet$ L/S 2.036 $\bullet$ L/O 2.485 $\bullet$ R/O 2.990 |
| **second NP** | | | | | | |
| # of fix. | — | — | — | 0.001 | 0.01 | R/S 2.780 $\circ$ L/S 2.813 $\prec$ L/O 3.677 $\circ$ R/O 4.069 |
| **image** | | | | | | |
| time | — | — | — | — | — | R/S 1946 $\wr$ L/S 2054 $\bullet$ R/O 2153 $\wr$ L/O 2479 |
| # of fix. | — | — | — | — | — | R/S 7.512 $\wr$ L/S 8.211 $\bullet$ R/O 8.363 $\wr$ L/O 9.414 |
| **agent** | | | | | | |
| # of fix. | — | — | — | 0.001 | 0.01 | R/S 1.524 $\circ$ L/S 1.711 $\prec$ R/O 2.578 $\circ$ L/O 2.879 |
| 1st fix. time | — | — | — | — | — | L/S 157.9 $\wr$ R/S 158.7 $\bullet$ L/O 173.1 $\wr$ R/O 211.7 |
| **O-CHAR** | | | | | | |
| # of fix. | — | — | — | 0.01 | 0.001 | R/S 2.304 $\circ$ L/S 2.319 $\prec$ R/O 3.353 $\circ$ L/O 3.808 |
| **left pic.** | | | | | | |

Continued on next page

Table 3.8 – continued from previous page

| Region / measure | ⇒≉⇐ both | ⇒≉⇐ SVO | ⇒≉⇐ OVS | SVO ≉ OVS ⇒ | SVO ≉ OVS ⇐ | Order of values |
|---|---|---|---|---|---|---|
| time | — | — | — | 0.001 | — | R/S 403.6 ≀ L/S 430.1 • L/O 529.2 • R/O 727.3 |
| # of fix. | — | — | — | 0.001 | — | R/S 1.524 ≀ L/S 1.675 • L/O 1.778 • R/O 2.578 |
| 1st pass time | 0.05 | — | — | 0.01 | — | L/S 241.9 ≀ R/S 263.5 • L/O 294.8 • R/O 446.2 |
| 1st pass # | 0.05 | — | 0.01 | 0.001 | — | L/S 0.976 ≀ R/S 1.012 ∘ L/O 1.020 ≺ R/O 1.578 |
| **middle pic.** | | | | | | |
| 1st fix. time | 0.05 | — | — | — | — | L/O 214.5 ≀ L/S 233.6 • R/O 244.2 ≀ R/S 261.2 |
| 1st pass # | 0.05 | — | — | 0.01 | 0.001 | R/O 1.382 • L/O 1.455 • R/S 1.744 • L/S 2.018 |
| **right pic.** | | | | | | |
| time | 0.01 | — | 0.001 | — | 0.001 | R/O 374.7 ≀ R/S 419.3 ∘ L/S 423.0 ≺ L/O 756.6 |
| # of fix. | 0.01 | — | 0.001 | — | 0.01 | R/O 1.520 ≀ R/S 1.673 ∘ L/S 1.711 ≺ L/O 2.879 |
| 1st pass time | 0.05 | — | 0.01 | — | 0.01 | R/O 229.9 ≀ L/S 262.7 ∘ R/S 265.8 ≺ L/O 422.5 |
| 1st pass # | 0.01 | — | 0.001 | — | 0.001 | R/O 0.941 ≀ R/S 1.054 ∘ L/S 1.084 ≺ L/O 1.657 |

fix.: fixation(s), #: number (of fixations), ins.: inspection, r-path: regression path
Underlined: significant effect in “nonstandard” direction (OVS ≺ SVO or ⇒≺⇐)

The table also shows some effect asymmetries: For example the sentence type effect on background fixations is only significant at $p < 0.01$ or better when the image action direction is to the left. This asymmetry is, however *not* a significant effect in itself. The table shows asymmetries of effects and post-hoc properties of the sorted averages only as supplemental information. Rows with symmetrical effects of sentence type or match / mismatch are not shown unless there is also a main effect or sub-effect of image action direction.

Only few regions apart from left / right character have significant main effects or sub-effects of image action direction. For the middle character, there are weak effects ($p < 0.05$) on first fixation time and number of first inspection fixations. The two effects follow different patterns and are in different directions. It is unknown how those effects can be explained. There is also a weak effect of image action direction on the NP1 first fixation time.

Table 3.9 shows the same overall effect of image action direction column as the previous table 3.8, but now places the data in context of interactions with match / mismatch: We now see a sub-effect of image action direction on S-CHAR first fixation time for match trials and a sub-effect of NP1 first fixation time for mismatch trials. We also see that the N-CHAR effects of match / mismatch are asymmetrical along the image action direction, but this asymmetry is no significant effect itself.

Table 3.9: Significance of to-right / to-left effects and left / right asymmetries in match (√) / mismatch (⋆) effects

| **Region / measure** | Is ⇒≉⇐? over-all | only for √ | only for ⋆ | match ≉ mismatch? significances for ⇒ | for ⇐ | **Averages**, sorted (R/Y is ⇒/match, L/N is ⇐/mismatch) |
|---|---|---|---|---|---|---|
| **first NP** | | | | | | |
| 1st fix. time | 0.05 | — | 0.01 | — | — | L/N 168.9 ≀ R/Y 184.5 • L/Y 186.6 ≀ R/N 201.3 |
| **S-CHAR** | | | | | | |
| time | — | — | — | — | — | R/N 841.2 ≀ L/N 943.4 • R/Y 1026 ≀ L/Y 1122 |
| # of fix. | — | — | — | — | — | R/N 3.103 ≀ L/N 3.357 • R/Y 3.522 ≀ L/Y 4.240 |
| 1st fix. time | — | 0.01 | — | — | — | L/Y 218.3 • L/N 256.0 • R/N 260.8 ≀ R/Y 262.5 |
| 1st pass # | — | — | — | — | — | R/N 1.603 ≀ L/N 1.800 • R/Y 1.940 ≀ L/Y 2.128 |
| **N-CHAR** | | | | | | |
| time | — | — | — | 0.01 | — | R/Y 166.0 • L/Y 218.4 • R/N 270.5 ≀ L/N 287.2 |
| # of fix. | — | — | — | — | — | R/Y 0.746 ≀ L/Y 0.856 • R/N 1.081 ≀ L/N 1.243 |
| 1st fix. time | — | — | — | 0.001 | — | R/Y 92.10 • L/Y 135.0 • L/N 139.6 ≀ R/N 157.7 |
| 1st pass time | — | — | — | 0.01 | — | R/Y 123.5 • L/Y 173.8 • L/N 192.4 ≀ R/N 198.0 |
| 1st pass # | — | — | — | — | — | R/Y 0.575 ≀ L/Y 0.704 • R/N 0.794 ≀ L/N 0.836 |
| **left pic.** | | | | | | |
| 1st pass time | 0.05 | — | — | — | — | L/N 258.5 ≀ L/Y 265.2 ≀ R/N 319.6 ≀ R/Y 345.7 |
| 1st pass # | 0.05 | — | — | — | — | L/N 0.971 ≀ L/Y 1.016 ≀ R/N 1.213 ≀ R/Y 1.239 |
| **middle pic.** | | | | | | |
| 1st fix. time | 0.05 | — | — | — | — | L/Y 219.9 ≀ L/N 232.4 • R/Y 245.4 ≀ R/N 264.1 |
| 1st pass # | 0.05 | — | — | — | — | R/Y 1.530 ≀ R/N 1.684 • L/Y 1.696 ≀ L/N 1.907 |
| **right pic.** | | | | | | |
| time | 0.01 | — | — | — | — | R/N 396.5 ≀ R/Y 408.5 ≀ L/N 493.3 ≀ L/Y 608.5 |
| # of fix. | 0.01 | — | — | — | — | R/N 1.507 ≀ R/Y 1.724 • L/N 2.007 ≀ L/Y 2.304 |
| 1st pass time | 0.05 | — | — | — | — | R/Y 249.7 ≀ R/N 254.7 ≀ L/N 298.5 ≀ L/Y 349.1 |
| 1st pass # | 0.01 | — | — | — | — | R/N 0.971 ≀ R/Y 1.052 • L/N 1.243 ≀ L/Y 1.360 |

fix.: fixation(s), #: number (of fixations), ins.: inspection, r-path: regression path

Underlined: significant effect in "nonstandard" direction (⋆ ≺ √ or ⇒≺⇐)

#### 3.2.7.5 Effects in *incorrectly* answered trials

As there were many incorrectly answered trials, in particular for the OVS case, we did a separate analysis of those trials. The results are shown in table 3.10 for the contrasts of SVO / OVS and match / mismatch. Data is quite sparse for SVO trials (only few answered wrong), so for the sub-effects, significance is often only reached for OVS. Tables with the

Table 3.10: Significance of SVO / OVS and match (√) / mismatch (⋆) effects and interactions of both in **incorrectly answered** trials

| **Region** / **measure** | Is SVO ≉ OVS? over-all | only for √ | only for ⋆ | match ≉ mismatch? over-all | only for SVO | only for OVS | **Averages**, sorted (S/Y is SVO/match, O/N is OVS/mismatch) |
|---|---|---|---|---|---|---|---|
| **background** | | | | | | | |
| 1st fix. time | 0.01 | — | — | — | — | — | O/N 90.75 ≀ O/Y 92.88 ≀ S/Y 140.8 ≀ S/N 160.7 |
| **first NP** | | | | | | | |
| time | — | — | — | 0.01 | — | — | S/Y 382.0 ≀ O/Y 498.8 ≀ S/N 566.8 ≀ O/N 622.5 |
| # of fix. | — | — | — | 0.01 | — | — | S/Y 2.095 ≀ O/Y 2.449 ≀ O/N 3.091 ≀ S/N 3.200 |
| 1st fix. time | 0.05 | — | — | — | — | — | S/N 145.3 ≀ S/Y 146.0 ≀ O/N 183.7 ≀ O/Y 196.0 |
| **verb** | | | | | | | |
| 1st fix. time | 0.01 | — | — | — | — | — | S/N 147.4 ≀ S/Y 176.2 • O/N 209.1 ≀ O/Y 228.1 |
| 1st pass time | 0.05 | — | — | — | — | — | S/N 178.4 ≀ S/Y 187.1 ≀ O/N 221.4 ≀ O/Y 256.7 |
| **patient** | | | | | | | |
| time | — | — | — | 0.001 | — | 0.001 | S/N 198.0 ≀ O/Y 331.1 ≀ S/Y 500.2 • O/N 786.3 |
| # of fix. | — | — | — | 0.001 | — | 0.001 | S/N 1.100 ≀ O/Y 1.205 ≀ S/Y 1.762 • O/N 2.831 |
| 1st fix. time | — | — | — | 0.05 | — | — | S/N 133.8 ≀ O/Y 191.1 ≀ S/Y 209.6 ≀ O/N 265.1 |
| 1st pass time | — | — | — | 0.05 | — | 0.01 | S/N 198.0 ≀ O/Y 236.5 ≀ S/Y 313.7 • O/N 363.0 |
| 1st pass # | — | — | — | 0.01 | — | 0.001 | O/Y 0.872 ≀ S/N 1.100 • S/Y 1.143 ≀ O/N 1.312 |
| **S-CHAR** | | | | | | | |
| time | 0.01 | 0.001 | — | — | — | — | O/Y 498.3 ≀ S/N 720.7 • O/N 786.3 ≀ S/Y 1307 |
| # of fix. | 0.01 | 0.001 | — | — | — | — | O/Y 1.885 ≀ S/N 2.800 • O/N 2.831 ≀ S/Y 4.571 |
| 1st fix. time | — | 0.01 | — | 0.01 | — | 0.001 | O/Y 164.6 • S/N 261.9 • O/N 265.1 ≀ S/Y 295.0 |
| 1st pass time | 0.01 | 0.01 | — | — | — | — | O/Y 313.9 ≀ O/N 363.0 • S/N 403.5 • S/Y 607.4 |
| **O-CHAR** | | | | | | | |
| time | 0.01 | — | — | 0.05 | — | — | S/Y 500.2 ≀ S/N 537.1 • O/Y 866.7 ≀ O/N 1146 |
| # of fix. | 0.01 | — | — | 0.05 | — | — | S/Y 1.762 ≀ S/N 2.100 • O/Y 3.090 ≀ O/N 3.922 |
| **N-CHAR** | | | | | | | |
| 1st fix. time | — | — | — | 0.01 | — | 0.01 | O/N 121.7 ≀ S/N 133.8 • S/Y 155.5 ≀ O/Y 191.1 |

fix.: fixation(s), #: number (of fixations), ins.: inspection, r-path: regression path
Underlined: significant effect in "nonstandard" direction (OVS ≺ SVO or ⋆ ≺ √)

analysis of the effects of image action direction in incorrectly answered trials can be found in the appendix – they are not discussed here[1].

Most of the effects found for correctly answered trials are completely absent for incorrectly answered trials: There is *no* overall effect of SVO / OVS or match / mismatch on the response times, for the image region group and the agent and ambiguous middle character regions. Interestingly, there are no effects on adverbial region or NP2 either. There is still an effect of match / mismatch on the patient region, which is mostly based on data for OVS trials. There are still effects of SVO / OVS sentence type for S-CHAR and O-CHAR regions. There are also first fixation effects of match / mismatch for S-CHAR / N-CHAR and "normal" (fixation count and duration) O-CHAR SVO / OVS and (weak) match / mismatch effects. There is an effect of match / mismatch for the first NP and an effect of sentence type on first NP first fixation time. There are first fixation effects of SVO / OVS on NP1 and verb region and a first pass time effect on the verb region.

The verb and first NP effects are *specific* to incorrectly answered trials: Those regions show *different* effects for correctly answered trials. Similarily, the patient region main effect is only visible for incorrectly answered trials: For correctly answered trials, two sub-effects of opposite direction on the patient were present but no overall main effect was present. The SVO specific sub-effect is absent from data for incorrectly answered trials. O-CHAR sub-effects are no longer significant. S-CHAR, N-CHAR and background effects are of different type and direction compared to the effects for correctly answered trials.

## 3.3 Discussion

### 3.3.1 Large scale effects

This chapter summarized the results of our analysis of the experiment logs. The data has given us some new insights on the way that our participants integrate simultaneously presented scene and written sentences. It turned out that all participants, when being asked to decide whether scene and sentence match, use the strategy of reading the whole sentence first, before inspecting the image closer. Those results fit with the observations of [RRS+01], who analyzed how people gather information from print advertisements and with [UJR04], who asked participants to do a yes / no decision about road scenes and simultaneously displayed sentences in one of their experiments.

The task for our participants was to decide about image / sentence match, with the stimuli containing both SVO and OVS sentences and images with action to the right as well as with action to the left. The OVS word order sentences, designed to be garden path style up to the final second noun phrase, had a striking impact on not only the eye movement pattern of the participants but also on the correctness of the answers: One third of the participants performed *worse than guessing* on the OVS trials, indicating that they just treated the OVS

[1]There are, apart from effects on left / right character, only effects on background (time, number of fixations) and S-CHAR (sub-effect of first inspection time for SVO) and some asymmetries of effects.

sentences as if they were SVO sentences. The others still had a much higher error rate for OVS sentences than for SVO sentences.

For 3 of the 32 participants, many trials could not be analyzed as there were extensive periods without eye-tracker data. For the others, however, average loss rate was about one out of the 24 item trials (not counting the filler trials) per participant. As eye-trackers generally can have problems to track the eyes of certain participants, a possible solution for future experiments can be to acquire a replacement participant (who gets the same trial script) whenever more than for example one third of the data is unusable for a particular participant.

### 3.3.2 Effects found in histogram and statistical data

We found many effects of sentence type in our data: There were longer inspections and more fixations for OVS trials than for SVO trials for many regions. The affected regions are O-CHAR, the sentence and image region groups, all four text regions, left and right character and the agent character. For the NP2 region, there is also an effect on the regression path measure. The main effects for NP1, verb and left and right character are also significant when we only check the match trials but not when we check for a mismatch trial sub-effect. The overall response time is longer for OVS trials. There is also a sub-effect of sentence type among match trials for S-CHAR first inspection data. The average per fixation duration is longer[1] for OVS trials than for SVO trials for the S-CHAR region.

As there were no first pass effects on any of the text regions, the overall inspection duration and fixation count effects can be attributed to re-reading. The strong effect on NP2 regression path backs this observation. There are first inspection effects for all image regions – normal because the participants read the text before inspecting the scene. Effects match our general predictions, but add some new details: For example we found that the text effects are related to re-reading. We also analyzed image fixation data using three different viewpoints.

There are also some regions for which SVO trials have longer inspections than OVS trials: Most importantly the patient region for match trials (as predicted). For mismatch trials, we get the opposite effect: The patient inspection is more thorough for OVS mismatch than for SVO mismatch. On the other hand, the agent inspection is more thorough for SVO mismatch than for OVS mismatch, and shorter for SVO match than for OVS match. We did predict an interaction effect for the agent and patient regions for the *match* trials: More fixations / longer inspection for the agent region for OVS and more fixations / longer inspection for the patient region for SVO.

Both the effects for the mismatch trials as well as the effects for the match trials on the agent and patient regions fit well with the observation that there are always the fewest fixations and shortest inspection times on the character region which shows the N-CHAR. We see in the EDA results that this contrast is significant for inspection time, fixation count, first fixation duration, first inspection time and first inspection fixation count for both agent and patient regions. The EDA also shows a significant contrast between “agent is O-CHAR” (SVO

[1]No statistical test done: Analyzing per fixation duration would require to take both in-trial and between-trial variances into account. Descriptively, sentence type has a clear effect here, but we did not measure the significance level.

mismatch) and "agent is S-CHAR" (OVS match) while there is no such significant (threshold: $p < 0.01$ for EDA) contrast for the patient region. In addition, measures for OVS trials have higher values than measures for SVO trials for the agent region but not for the patient region.

Other cases where measures are higher for SVO trials than for OVS trials are middle ambiguous character region first inspection time and several S-CHAR measures, but the latter only show this contrast for the mismatch trials. The affected S-CHAR measures are overall and first inspection duration and fixation count. Note that S-CHAR first inspection measures show a contrast in opposite direction for match trials. The N-CHAR region data does not show any effects of sentence type at all.

The agent / patient effects of SVO / OVS sentence type for match trials are consistent with our predictions. Interestingly, we found that agent / patient sentence type effects of opposite direction are present for the mismatch trials. We did not predict the middle character effects nor the S-CHAR mismatch effects of sentence type.

There were fewer effects of match / mismatch than expected: We assumed that a matching scene would faciliate the comprehension of OVS sentences, but actually reaction times only got shorter for SVO match trials. Decision was significantly faster for SVO match than for any other condition, while the other conditions did not differ from each other significantly. This means that decision was faster for SVO than for OVS, faster for SVO match than for OVS match, and faster for SVO match than for SVO mismatch. The contrast between SVO match and the other conditions is large, so there is a main effect of sentence type: Decision is significantly faster for SVO than for OVS. However, there is no main effect of congruence: Decision is *not* significantly faster for match than for mismatch trials in general.

We found a similar pattern for the NP2 regression path data: shortest inspection / fewest fixations for SVO match, longer / more for all other cases. However, even SVO mismatch has shorter inspection / fewer fixations than OVS match and OVS mismatch here. This means there is a match / mismatch effect limited to SVO sentences *and* an overall SVO / OVS effect for the NP2 regression path measures.

While we found *no* match / mismatch effects on the sentence or image region groups, we did find effects on N-CHAR and (limited to the first inspection measures) on O-CHAR. For the S-CHAR region, there are actually more fixations / longer inspections (overall and first inspection) for match than for mismatch. The EDA shows that this is because the OVS mismatch case has fewer / shorter fixations on S-CHAR than any other condition. Values are actually lower than those for O-CHAR here. The participants exhibited an inspection pattern which suggests that they treated S-CHAR and O-CHAR similar to how they treated O-CHAR and S-CHAR (respectively) for the other conditions.

Another observation which supports this assumption is that all conditions *except* OVS mismatch show a pattern of S-CHAR > O-CHAR > N-CHAR. In other words, SVO normally has NP1-CHAR > NP2-CHAR > N-CHAR and OVS normally has NP2-CHAR > NP1-CHAR > N-CHAR. Again, for OVS mismatch, data does not fit with the rule. The observations can be summarized as OVS mismatch being treated more like a SVO trial than like an OVS trial by the participants in their image inspection pattern. One could conclude that they came to their correct NO answer because "the NP2 determiner was wrong" rather than "the depicted action did not fit the sentence", but it would take further experiments to investigate this.

As announced earlier in this thesis, we now describe the inspection contrasts on the image regions with some rules. The minimal set of rules is: 1. Shortest inspection / fewest fixations are on N-CHAR. 2. There are more fixations / longer inspections of the middle character than of the patient character. 3. There are more fixations / longer inspections of S-CHAR than of O-CHAR *unless* this conflicts with rule 2 (which is the case for OVS mismatch trials).

If we base the rules only on the agent / middle / patient distinction, we get: 1. SVO match and OVS mismatch have patient > agent, while OVS match and SVO mismatch have agent > patient. For the match trials, this is exactly what we have predicted. 2. middle > patient. 3. middle > agent *except* for OVS match. The mismatch case of rule 1 can be interpreted as: The participants looked at the mentioned characters, even though they have their roles swapped (but still take part in the mentioned activity). Rule 3 could also be interpreted as "... except for OVS except for OVS mismatch" to emphasize the expected effects of sentence type on agent / patient inspection ratio.

As mentioned above, we can also describe the data in terms of S-CHAR, O-CHAR and N-CHAR: S-CHAR > O-CHAR except for OVS mismatch. Always S-CHAR > N-CHAR and O-CHAR > N-CHAR. The OVS mismatch exception also is visible if we use NP1-CHAR, NP2-CHAR and N-CHAR as labels, as seen above.

Finally some patterns which are only visible in our per-condition histograms: For SVO, we have more S-CHAR > O-CHAR and only little agent / patient differences. The S-CHAR peak is before the O-CHAR peak. For OVS, we have more agent / patient differences and only little S-CHAR / O-CHAR differences. The O-CHAR peak is before the S-CHAR peak. Actually, there are initially more fixations on O-CHAR than on S-CHAR and later more fixations on S-CHAR than on O-CHAR. A higher percentage of fixations is on NP2 for OVS compared to SVO in many slots. There are also more fixations on NP2 for the mismatch trials, and less image fixations. For the match trials, however, the peak of NP2 fixations is at a later slot and has a higher level (percentage of all fixations at that slot) than for the mismatch trials. The S-CHAR > O-CHAR difference is bigger for the match trials.

There seems to be some bias to look more at depicted active / agent characters than at depicted passive / patient characters, too, but the effect of the sentence role seems to be stronger than the effect of the image role.

There was a slight effect of image action direction on some middle character and first NP measures. The histograms give no explanation for that effect, but manually looking at the visualized fixation logs with DPIFilter suggests that it could be related to a possible slight preference for a "circular" inspection pattern: Forced start in the middle, read the text on the top from left to right, inspect the characters at the bottom from right to left.

As there were many incorrectly answered trials, in particular for OVS stimuli, we also checked for effects in the data for incorrectly answered trials: The main finding here is the absence of many of the effects which are found in the correctly answered trials. However, one effect grew stronger compared to the correctly answered trials: There are fewer fixations on the S-CHAR in the OVS match case, probably because the participants treated it as the O-CHAR (mistook the sentence for a SVO one) on their way to giving the wrong answer to the decision task.

# Chapter 4

# General Discussion

## 4.1 Software performance

The NEWTRACK software got used for the first time in a complete eye-tracking study. We have been able to replicate established effects of congruence and sentence type with it. Some informal statistical analysis of our logs supports the assumption that the recorded fixations (after filtering and pooling) have similar properties as those recorded by, for example, our READING software. With both systems, about 85% of the trial time is part of fixations, gaps between fixations are short, and the histograms of fixation duration distribution are similar.

NEWTRACK performed well on near-standard hardware, a 700 MHz PC with two graphics cards and a custom digitizer box. The latter could have been replaced by an industrial A/D board, but it was felt that there was no need for higher resolution or speed: The custom box already reaches, depending on how much of the tracking range is actually used, up to several 1000 steps per axis of resolution, and up to several 1000 samples per second. In addition, the custom box is not limited to a specific bus system / architecture: It attaches to the printer port externally. Note that using USB to printer port converters is *not* supported here, as those would reduce access speed far too much.

Because all the low-level work of eye-tracking is done by the hardware of a Fourward DPI eye-tracker, NEWTRACK focuses on interactive stimulus presentation and on high-speed data processing. Presentation and processing are working well for the concurrent scene and sentence presentation in our study.

The data processing has still some room for improvement: For example configurable fixation onset and offset eye movement thresholds and configurable minimal fixation times, along with a built-in fixation pooling function, would lead to smaller and more directly useable log files. On the other hand, the current NEWTRACK log files already include better support for later analysis than for example the logs generated by READING, our SMI Eyelink based software. And it can sometimes be useful to have more verbose logs, although there was no need for more verbosity in the current study. Another very useful feature of NEWTRACK is that it can convert stimulus images into templates (images where each depicted item is represented as a flat colored blob) quickly and automatically for later analysis of the fixation locations.

Cleaning up of the logged data was done with the custom DPIFilter software. DPIFilter is a modification of the AscFilter software for READING. After some initial problems to get it to run stable, DPIFilter has proven really useful, and got enhanced with some new features which are to be backported to AscFilter. While the plain data filtering and pooling component is just a necessity, the data visualization and correction part has been very useful as well, in particular in conjunction with NEWTRACK: The visualization allows direct human judgement whether the logged data is plausible, and it gives a direct visual impression of the eye movements of the participants, telling the human operator a lot which he would not be able to realize from just reading the logged coordinates or doing statistics on them.

In addition, the correction interface helped to cope with a weakness in NEWTRACK: The used version had no drift correction, and as it was not feasible to recalibrate the eye-tracker often enough to give exact coordinate offsets without disturbing the participant too much, the manual post-correction in DPIFilter was just the right solution.

### 4.1.1 Drift correction

Recent versions of NEWTRACK add a drift correction feature, which was not available in the version used for the experiment described in this thesis. The section 3.1.4.2 analyzes the amount of required offset corrections for our experiment and suggests to use drift correction to reduce the required corrections. Drift correction mode is similar to a “gaze trigger” mode.

In drift correction mode, the gaze mark shown at the start of a trial is not only taken as fixation target for manual verification of the tracking. Instead, very long fixations exactly at the gaze target directly trigger the display of the stimulus. In addition, the operator can still manually trigger the stimulus display. This second trigger method has a new side effect in drift correction mode: If the gaze location is within a certain range of the expected location, the software assumes that the gaze was *exactly* at the expected location, and corrects the coordinate system offset accordingly.

The current drift correction limits of the updated NEWTRACK software allow a (total, not relative) coordinate system shift of up to 2.5 character sizes sideways and up to 1.5 character sizes vertically, based on the current $16 \times 32$ font size. This corresponds to correction vectors ranging from $(-40, -48)$ to $(40, 48)$. Experience will show if this range should be reduced. It might be useful to automatically enter calibration mode when the range is exceeded. The current implementation just discards drift correction attemps that would be out of range. Drift correction results are logged, storing the correction vector between most recent calibration and the coordinate system after the drift correction.

The current version of NEWTRACK will need less DPIFilter post-corrections, as drift correction has been added to NEWTRACK recently. Still, DPIFilter in itself already is an improvement over AscFilter, having the ability to assign nearby (radius of about 5 to 15 pixels recommended) fixations to image items. This margin processing had only been possible for text items in AscFilter.

### 4.1.2 Suggested future NEWTRACK improvements

As written earlier in section 4.1, it would be useful to have configurable fixation onset and offset sensitivity in NEWTRACK. The current setting is asymmetrical, modelled after recommendations for very sensitive fixation / saccade detection (e.g. useable in gaze-contingent context), which can lead to some very short "fixations" being logged around the core of a fixation. Those short fixations are pooled into the main one by DPIFilter's pooling function anyway, but avoiding them in the first place would make the raw logs better to read for humans.

Calibration is using a $3 \times 3$ grid at the moment, selecting two best out of three measurements for each data point. A future version could estimate whether all three values match reasonably well, and if not, require resampling of bad calibration datapoints. Another possible enhancement for calibration would be to use more gaze targets, to fit a more advanced transformation from measured raw coordinates to actual screen coordinates, for example allowing tilted screen placement or highly nonlinear transformations.

The Fourward DPI tracker has good linearity and the fixed lab setup allows to avoid effects like distortions caused by asymmetrical screen placement or using a screen with curved surface. Linearity issues with the A/D hardware can be avoided even with low cost digitizers, too: Those factors combined suggest that, having the system calibrated for a single participant in a fixed lab environment, calibration for other participants could use stored properties of the master calibration, adding only some minor corrections customized for the particular participant. This would speed up the (re-)calibration process during experiments.

There are several drivers which could be useful additions for NEWTRACK, such as support for systems with a single dual-head / dual-core graphics card instead of two standard graphics cards. Another driver issue is support for some standard data acquisition boards, for those who want extra speed and precision or want to avoid building their own A/D box. Support of the old ISA DAS16 is already included, but PCs with ISA slots are rare nowadays outside the realm of special industry systems – or museums.

One driver which would be really interesting to have for NEWTRACK is support for "SoundBlaster" or "AC97" compatible soundcards and on-board sound systems: This would allow us to run experiments with different modalities (spoken versus written text, for example) on the same system. Experimental results could be compared more directly that way, as more factors are held constant when using the same software for both visual and audiovisual experiments.

## 4.2 Our results in context of other studies

Our findings from the concurrent scene and sentence presentation study include that the participants first read the sentence (as in [UJR04] and [RRS+01]), possibly looking twice at ambiguous parts, before they do a closer inspection of the scene. The scene inspection is controlled more by the text contents than by the image contents: Earlier and more looks go to the S-CHAR region, followed by the O-CHAR region. There are only a few looks to the N-CHAR region. Guidance by scene rather than by text contents could have meant more contrast between agent and patient regions rather than between S-CHAR and O-CHAR. Judging from our data, this would mean that the scene contents have more relative influence on the inspection pattern for OVS trials and less influence for the SVO trials.

We found evidence for various strategies of the participants after reading the sentence and looking at the scene, in particular in the alledgedly harder mismatch and OVS cases: Re-reading the ambiguous part of the text, re-reading the whole text, and closer image inspection.

We found longer reading times on all text regions for OVS trials than for SVO trials: This is consistent with the [KC05] findings, but our participants read the text *before* inspecting the image. The serial presentation in [KC05] forced the participants to inspect the image before reading the text. Our participants selected the wrong answer in the match / mismatch decision task very often for OVS trials. We found that the longer reading times of OVS trials are caused by re-reading rather than by longer initial (first pass) inspections.

The [KC05] results show similar but not identical effect patterns compared to ours: In both studies, the sentence type factor has significant effect on each of the text regions. There are only few effects of congruence. In [KC05], there are effects on verb and adverbial region. In our results, there are no such effects. Instead, there are SVO-only effects of congruence limited to the NP2 regression path[1] data and the overall sentence inspection time. The latter is interesting as neither our study nor the serial presentation experiment of [KC05] found a main effect (not limited to SVO) of congruence on overall sentence inspection time. On the other hand, [UJR04] found an effect only for the concurrent presentation version of their experiment but not for serial presentation. As [UJR04] use only simple declarative sentences, their experiment can be compared with the SVO subset of our stimuli, and our study replicates the findings of their concurrent presentation experiment.

Comparing the three studies further, we find that only concurrent presentation shows overall sentence inspection time effects. We also find that congruence effects are small for OVS stimuli. While [KC05] did find a main effect of congruence on two text regions for serial presentation, their diagrams show that effects are much weaker for the OVS part than for the SVO part of their data. This contrast is even stronger for our results, where $p < 0.01$ sub-effects of congruence show up *only* for SVO for some measures: Answer time, overall inspection time, sentence inspection time and the NP2 regression path data. For all of those, neither overall main effects nor OVS sub-effects are found (all $p > 0.01$; the corrected threshold for the post-hoc EDA / planned comparisons is 0.01).

It is interesting to see that the serial (image first) presentation actually *introduced* / enabled congruence effects on verb and adverbial region. This contrasts with the fact that it *"disabled"* (or at least reduced below significance) congruence effects on the overall sentence inspection time.

The image inspection patterns of the visual world / spoken utterance experiment of [KCSP05] could be succesfully reproduced in our concurrent text and scene presentation experiment: More looks to the patient for SVO match and more looks to the agent for OVS match. So

[1]For NP2, this is defined as the set of fixations on any of the text regions starting with the first fixation on the NP2 region.

their results hold across modalities (spoken versus written text). The [KCSP05] study does not investigate congruence efffects, so they use only match trials.

As our participants read the whole text before looking at the image, we could not detect incremental effects. For example the verb of the sentence already selects an action before the NP2 determiner disambiguates the sentence structure. Yet, most participants only inspected the image after reading verb and NP2, not already after reading the verb. For concurrent presentation of single-channel (no audio, just visual) stimuli, this limitation will be hard to circumvent. Only for longer texts, [Heg92] found intermediate looks at the image while participants were reading the text.

In the mismatch trials, our participants looked at both mentioned characters, being involved in the correct activity, even though they had their roles swapped with respect to that activity. They only rarely looked at the character which is in the correct (matching with the text) role relation with the middle ambiguous (mentioned by NP1) character instead.

As seen in [KCSP05] and [TSKES95], distractor items on the scene only get very few fixations. Even though the three characters in our scenes are all equally likely to be involved in the action mentioned in the text, we still get a very few fixations on the N-CHAR region. This interacts with our agent / patient measurements, but we were still able to get some interesting results about those regions. It is possible that OVS mismatch stimuli were treated as “SVO with the wrong NP2 determiner” by our participants. In that case, the scene inspection would have lead to a decision about the sentence structure. This does not seem to happen for the SVO mismatch and the match trials: For the latter, the text is read and the scene is found to match. For the former, as SVO is the preferred word order in German, the scene is probably judged as mismatching without attempting to reinterpret the sentence as “mistyped OVS sentence”. Further experiments could investigate that.

The authors of [KSA03] see a strong preference for SVO (or: agent first) interpretation as the explanation of some agent / patient style asymmetry: In their experiments with German SVO / OVS and English active / passive sentences, one region showed a much stronger effect of sentence type than the other. With (spoken) sentences like “The hare will be eaten by the fox” and “The hare will eat the fox”, the sentence type hardly influenced the amount of fixations on the cabbage (patient). However, there was a main effect on the amount of fixations on the fox (agent). The authors do not report whether there were effects on the ambiguous character (the hare) in the scene.

Our results are similar in that we found stronger effects (main effects and properties of the ordered-values EDA analysis results) of sentence type and role on the agent region than on the agent region. Note that [KSA03] did not manipulate match / mismatch, their experiment used a “look and listen” task. Their German experiment (NP1 case marking selects either SVO or OVS) showed more but still not statistically significant effects on the patient area compared to their English active / passive experiment.

In an experiment with spoken utterances, an interesting conflict between role and character reference could arise: The NP2 determiner selects the role of the character which will be mentioned by the NP2 noun. The determiner also indirectly selects the role of the NP1-CHAR. A moment later, the NP2 noun selects the role filler for the NP2. The distance between NP2 determiner and NP2 noun can be stretched (e.g. “Die Putzfrau malt der *bärtige* Matrose”, The cleaning lady[nom/acc] paints the bearded sailor[nom] / The bearded[1] sailor paints the cleaning lady).

Not all predicted match / mismatch effects were found: While a mismatch caused longer decision times for SVO, this was not the case for OVS. Decision times were generally long for OVS, and even insignificantly *longer* for OVS match than for OVS mismatch. In that, results are similar to the [KC05] results. To fit our data with the model of [CJ75], we would have to assume that the determiner which disambiguates the sentence to OVS by marking the NP2 as nominative is a "negative polarity marker". In other words, the case marker marks the sentence as "not SVO" and SVO is the default sentence structure. We have matching constituents but mismatching roles for the predicate / action in our mismatch trials: The OVS case marker would trigger recoding of the action before scene representation and sentence representation are compared. In that comparison, the mismatch case causes an extra delay. So the model predicts OVS mismatch > OVS match > SVO mismatch > SVO match reaction times, which does not fit with our reaction time data.

The Carpenter and Just prediction *does fit* with our sentence inspection duration data, but the OVS match / OVS mismatch difference is not significant in our results. O-CHAR, adverb and NP2 data shows the predicted ordering as well, but again without significant contrast between OVS match and OVS mismatch. We have strong effects of OVS being harder than SVO, but effects of congruence seem to be mostly limited to SVO. As most of our scene contents are in a complex relation to congruence and sentence type, we cannot compare our data to [UJR04] directly: Their study found bigger congruence effects on the text than on the image. In our experiments, the sentence type factor distorts possible effects of congruence on image fixations. Only for the N-CHAR region we can report that the effect of congruence has similar strength than the effect of congruence on the overall text inspection time. Note that both effects are not significant when looking only at data from OVS trials.

## 4.3 Experiments with object-first sentences in German

While we had expected OVS sentences to induce more errors in the match / mismatch decision task, results turned out to be almost too extreme. A whole group of participants had treated most of the OVS sentences as if they were SVO sentences. A factor here is the initial ambiguity of the sentences: Only the determiner of the second noun phrase disambiguates the sentence as SVO or OVS, after the reader has followed the default SVO assumption for a long time. Several participants therefore either overlooked the case marking towards OVS or intentionally denied it. There were actually some participants claiming (after the experiment) that the stimuli contained "spelling errors" in the determiners.

The remaining group of participants had, as expected, an elevated number of misdecisions for the OVS trials, but only a few of them getting close to the error rate that could also be achieved by mere guessing. So our experiment had the expected result for the majority of the participants – but still, having one third of "OVS agnostics" among our participants, gives a clear sign of the depreciation of this word order in everyday German language. Having clearer

[1] Of course the "bearded" must not disambiguate the reference in this case, so the scene for the sentence would show two bearded people, one of them being a sailor.

OVS markers, e.g. both subject and object noun phrase with unambiguous case markers, or the depicted activities having a strong bias towards one subject / object role assignment, would of course still make it possible for a bigger percentage of participants to understand OVS right. However, such extreme measures to point the participants to the fact that they are processing OVS sentences are often no option in psycholinguistic experiments like the one described in this thesis.

# Chapter 5

# Conclusions

Our new NEWTRACK eye-tracking software and the DPIFilter log-cleaning and visualization software have been shown to work well. We have been able to find expected established fixation effects and get new insights using the software. With NEWTRACK, a standard PC with a second graphics card and some simple add-on data acquisition hardware can provide a full eye-tracking environment, while post-processing can be done in a platform-independent way with the Java-based DPIFilter and the portable GNU C image chunking component of NEWTRACK.

Some suggested improvements for NEWTRACK would change some aspects of the raw log-files, further reducing the necessary post-processing work. Additional drivers can make NEWTRACK work on a wider range of hardware, or support new functions like audio playback. NEWTRACK is open source software, written in the DOS variant of 32bit GNU C. Overall system design is well-documented and encourages additions by other users, helping to improve the system for all users. Ballast like reliance on ancient hardware, limitations of 16bit compilers or having numerous unmaintained old code fragments in the source code has been avoided. NEWTRACK has been written from scratch, with well-sectioned and well-documented sources, standard hardware in mind.

We found that our participants read the sentence before doing closer scene inspection. This is in line with earlier studies. It biased the eye movement patterns towards text guiding image inspection rather than the other way round. However, it seems that the influence of the text on image inspection is weaker for OVS trials.

As others, we found strong effects of sentence type, related to re-reading rather than initial reading. We only found limited effects of congruence. In particular, effects are mostly limited to the SVO subset of our stimuli. Comparing our study to others, we found that serial presentation leads to different text inspection effects. We have been able to reproduce results of [KCSP05] in our experiment, across modalities. We did not notice incremental sentence processing effects, though, mostly because the written sentences were usually read completely before inspecting the scene. We presented some parallels between our results and those of [KC05] and [UJR04].

Some of our results fit well with the [CJ75] model, while other results are somewhat hard to explain in the lines of that model. In particular, effects of OVS sentence type and the interaction between congruence and sentence type do not fit well with [CJ75]'s model. Further research would be needed to investigate the differences and find ways to adjust the model to details of our data.

We found it useful to analyze the image inspection data from not only the agent / patient viewpoint but also from the S-CHAR / O-CHAR / N-CHAR viewpoint. By looking at scene regions by their text-determined role, we got additional insights beyond the standard sentence type / visual role interaction style analysis. The generally very detailed analysis of our data invites comparison to results of other experiments.

The OVS mismatch trial data exhibits several properties which seem to suggest that the participants interpreted the stimuli as misspelled SVO rather than as OVS with mismatching scene. Further experiments could shed some more light on that observation. Preference for SVO seems to be very strong: For example [KSA03] suggest that a strong default for SVO biased their results. Our experiment results contained a striking amount of incorrectly answered trials for OVS. At closer inspection, one third of our participants were found to have consistently selected SVO style answers for all or almost all OVS stimuli presented to them.

A possible follow-up experiment could use stronger hints to mark the OVS cases as such, to reduce the amount of incorrectly answered trials there. However, this would reduce possibly interesting "sentence type garden path" effects and limit our ability to put carefully designed ambiguities in our stimuli. Variation in the association between middle image and first noun phrase should not be needed for this modified experiment. The fixed association keeps the experiment design at a reasonable count of conditions, so this property of the stimuli should be preserved. Yet another possible variation of our experiment is to use images with a strong plausibility bias or with some stereotypical activities shown: Increasing the "weight" of depicted actions could interact in interesting ways with what we found to be common strategies, like using mainly the text as guidance for the image inspection pattern.

# Appendix A

# Materials

## A.1 Instructions for the participants

The following instructions were shown to the participants prior to the experiment. A translation for the highlighted sections is provided below. The actual instructions did not show any highlighting. The participants were given the instructions on a sheet of paper.

Spoken instructions would be more likely to prime the participants towards one direction of match-checking, e.g. looking at the text and then comparing the image to the actions described in the text. As the gaze mark which is shown between trials is at the center of the screen, the participants are forced to start each trial by looking at the image for at least a moment.

Still, all participants decided to read the whole sentence after only a short glance at the image, and only did more in-depth inspection of the image *after* reading the sentence once.

Psycholinguistisches Experiment
Anleitung und Informationen

**Vielen Dank für deine Teilnahme an diesem Eyetracking-Experiment. Bei dem Experiment werden mit einem sogenannten Eyetracker deine Blickbewegungen zur späteren Auswertung aufgezeichnet.**

**Das Gerät muss zu Beginn des Experimentes auf dich justiert werden:** Bitte setz dich nahe an den Tisch, das Kinn auf die Kinnstütze gestützt und die Stirn gegen die oberen Stützen gelehnt. Achte darauf, bequem zu sitzen, und den Kopf während des Experimentes ruhig zu halten, um Messfehler zu vermeiden.

Wichtig: Klopfe auf den Tisch und warte, bis der Eyetracker auf Pause geschaltet ist, bevor du sprichst oder aufstehst. Anders kann der Eyetracker bei dem Versuch, den Bewegungen zu folgen, ernsthaft mechanisch beschädigt werden. „Murmeln" ist ungefährlich, aber führt zu Messfehlern. Also erst klopfen.

Die Messung verwendet einen schwachen unsichtbaren Infrarot-Lichtstrahl und funktioniert auch mit Kontaktlinsen, leider aber nicht mit Brillen. Zunächst wird der Strahl mit Hilfe einer Infrarot-Kamera grob positioniert. Für die erste Feineinstellung musst du einen Punkt in der Bildmitte fixieren.

Zur Eichung werden dann nacheinander an neun Stellen auf dem Bildschirm blinkende Markierungen angezeigt. Schau bitte jeweils genau auf die Markierungen, bis sie verschwinden. Sobald alle Einstellungen gemacht sind, kannst du ausprobieren, ob der PC deinen Blickbewegungen korrekt folgt.

**Im Experiment bekommst du Bilder und Texte (jeweils gleichzeitig) zu sehen, und musst entscheiden, ob Text und Bild zusammenpassen. Das Experiment dauert etwa eine halbe Stunde. Die oben beschriebene Kalibrierung (Eichung) kann dabei nach Bedarf zwischendurch wiederholt werden.**

**Zur Übung wird vor dem eigentlichen Experiment ein kurzer Durchlauf mit einigen Beispiel-Sätzen und -Bildern gemacht. Dann hast du nochmal die Gelegenheit, Fragen zu stellen. Für das eigentliche Experiment wird der Eyetracker dann wieder neu justiert.**

**Vor jedem Bild / Satz erscheint ein leeres Bild mit einer Markierung. Sobald du Satz und Bild sehen willst, schaue bitte genau auf die Markierung. Der Versuchsleiter kann so überprüfen ob die Erfassung noch gut ist, und wird die Anzeige aufdecken.**

**Schau dir die Anzeige an und entscheide dich, ob Bild und Satz zusammenpassen. Drücke dann sofort die entsprechende Antworttaste (JA oder NEIN). Die Anzeige wird dann wieder verdeckt.**

Wenn du sprechen oder das Experiment unterbrechen willst, klopfe zuerst auf den Tisch und warte, bis der Eyetracker auf Pause geschaltet ist, um Schäden zu vermeiden.

Für Fragen stehen wir natürlich gerne zur Verfügung. Vielen Dank für deine Teilnahme!

### Translation of key sections of the participant instructions

*The following translation tries to compromise between staying close to the original wording and keeping sensible language style:*

Thank you for participating in our eye-tracking study. During this experiment we will record your eye movements with a so-called eye-tracker for later analysis.

The device first has to be adjusted for you...

*(Some information about the used technology and about how to keep the eye-tracker happy follows, along with instructions for the initial setup and calibration. The participant will see a gaze-controlled cursor and a test grid after the calibration. This provides direct feedback about whether the calibration went right and a good way to show the participant which data is available to the machine...)*

During the experiment you are shown images and texts (simultaneously) and have to decide whether text and image match. The experiment takes about half an hour. The calibration which is described above can be repeated during the experiment as needed.

To practice, a short run with some example sentences and images is done before the actual experiment. You then have the chance to ask questions. The eye-tracker will be calibrated again for the actual experiment.

Before each image / sentence, a blank image with a mark is shown. As soon as you want to see sentence and image, look exactly at the mark. This allows the operator to verify if the tracking is still good, and he will then reveal the display.

Look at the display and decide whether image and sentence match. Then press the corresponding answer button (YES or NO) at once. Then the display is covered (hidden) again.

*(Concluding remarks follow: The participant is reminded to wait for an okay from the operator before leaving the tracker for any reason, questions are welcome, and we thank the participant for helping with our research.)*

## A.2 The practice items

The script with the practice items is shown in table A.1. It also illustrates the syntax used in the experiment scripts:

First, properties for the trial categories "Item" and "Filler" are defined. Both start with showing a drift correction mark and both allow the left and the right button as answer. Only for "Item" trials the fixation data is logged.

Then the trials are defined, one line defines one trial. The columns are, from left to right: Trial category, trial-id (which is logged), timeout (trials end after 60 seconds if the participant did not press any button), "image2" to specify a trial with 1 line of text and 1 image, the distance between screen and image (here: 5 pixels from the bottom), the image file name and (all further columns) the text itself.

## A.3 The experiment items

This section lists all 24 item groups of the eye-tracking study: Each group consists of two images (which differ in the action direction) and two SVO and two OVS sentences. The experiment design is orthogonal, so for each image there is one SVO and one OVS sentence which matches the image. The mismatch variants of each stimulus item are generated by flipping the assignment between images and sentences.

Table A.2 shows one item group in detail. At the end of this section, all trial groups are shown in condensed form.

Table A.1: The script for presenting the practice trials, an example of the NEWTRACK scripting language.

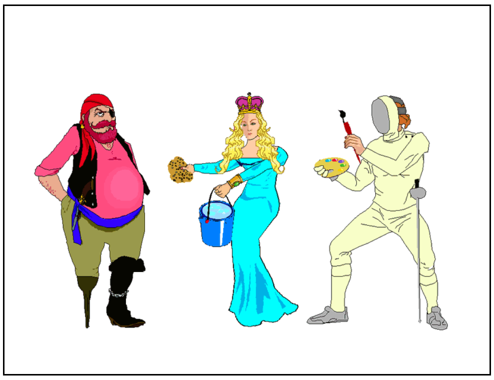
```
# Sechs Beispiel-Items, um zu demonstrieren, wie das Experiment abläuft.
define Item drift stream left right
define Filler drift nostream left right

Filler p1-m 60000 image2 -5 p1.pcx Der alte Mann bringt dem Pianisten die Noten.
Filler p2-n 60000 image2 -5 p2.pcx Die Reporterin gibt dem Zwerg ein Glas Wasser.
Filler p3-n 60000 image2 -5 p3.pcx Die Bardame bedient der Medizinmann.

Item p4-m 60000 image2 -5 p4.pcx Auf dem braunen Schreibtisch liegt ein Buch.
Item p5-n 60000 image2 -5 p5.pcx Der Soldat gibt dem Taucher eine blaue Decke.
Item p6-m 60000 image2 -5 p6.pcx Die Bardame erhält von dem Medizinmann Geld.
```

Each participant only sees one trial of each group. Trials versions are distributed across a set of eight experiment scripts, based on a Latin Square (see [@WOLFRAM]) design. As there are 24 (a multiple of 8) trial item groups, all participants are shown exactly three items of each of the two balanced variants (left / right) of the four conditions.

Experiment scripts are generated in sets of eight, so the number of participants has to be a multiple of eight. The order of the trials is pseudo-random: Between each two actual experiment items, one, two or three of the filler items are inserted (one random filler item is always inserted, then the remaining fillers are randomly placed between items which do not already have three fillers between them). The relative order of the experiment items, as well as the order of the filler items, are shuffled a bit at random, separately for each participant, during the generation of the scripts.

Table A.2: Example experiment item group: Eight orthogonal variants for each pair of images. Two sentence types, two match / mismatch truth values, and, for balancing, the two action directions.

| match | mismatch | Type | Text |
|---|---|---|---|
| ⇐ | ⇒ | SVO | Die Prinzessin wäscht offensichtlich den Piraten. |
| ⇐ | ⇒ | OVS | Die Prinzessin malt offensichtlich der Fechter. |
| ⇒ | ⇐ | SVO | Die Prinzessin malt offensichtlich den Fechter. |
| ⇒ | ⇐ | OVS | Die Prinzessin wäscht offensichtlich der Pirat. |



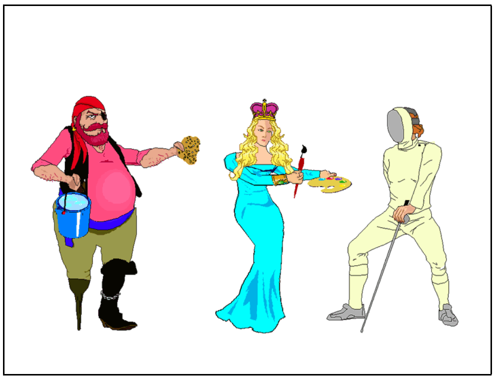
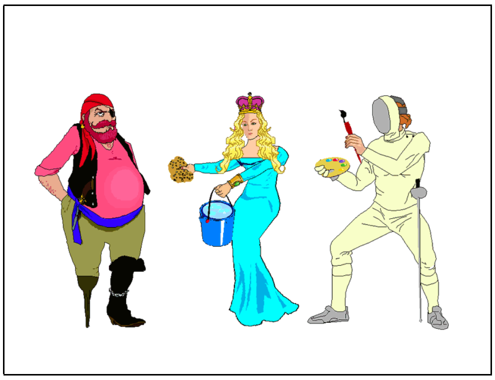
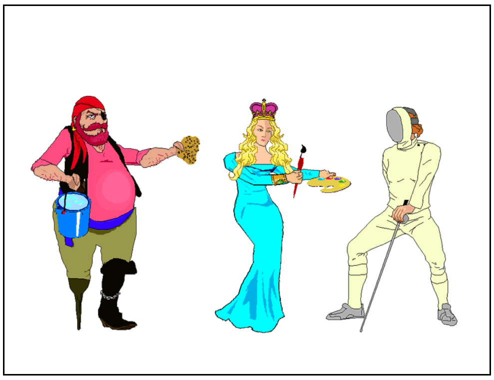


Table A.3: List of all used stimuli

Die Prinzessin. . .
. . . wäscht offensichtlich den Piraten.
. . . malt offensichtlich der Fechter.



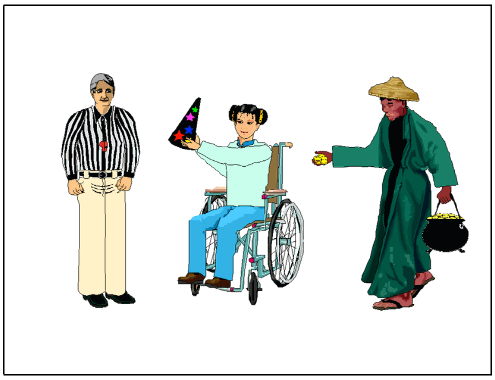
Die Prinzessin. . .
. . . malt offensichtlich den Fechter.
. . . wäscht offensichtlich der Pirat.



Die Badenixe. . .
. . . maskiert soeben den Skifahrer.
. . . entlohnt soeben der Musketier.

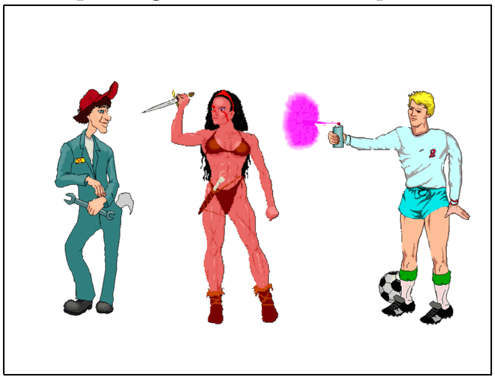
Die Badenixe. . .
. . . entlohnt soeben den Musketier.
. . . maskiert soeben der Skifahrer.

Die Rollstuhlfahrerin. . .
. . . kostümiert hier den Schiedsrichter.
. . . besoldet hier der Chinese.

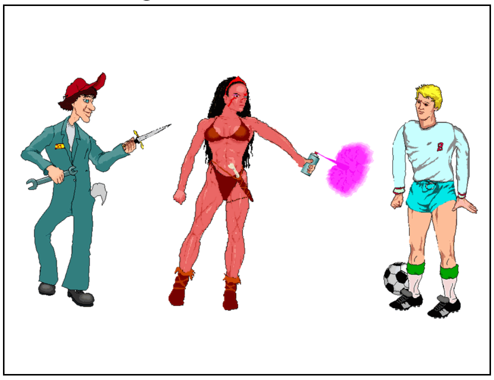


Die Rollstuhlfahrerin. . .
. . . besoldet hier den Chinesen.
. . . kostümiert hier der Schiedsrichter.

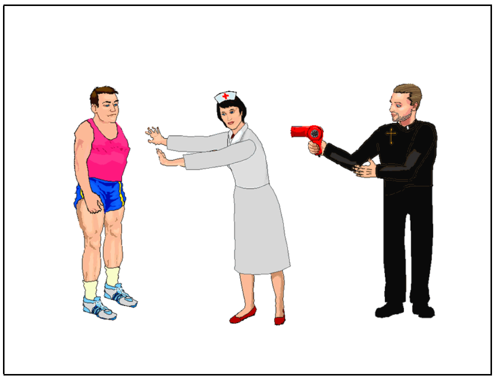
Die Amazone. . .
. . . erdolcht gerade den Mechaniker.
. . . besprüht gerade der Fußballspieler.



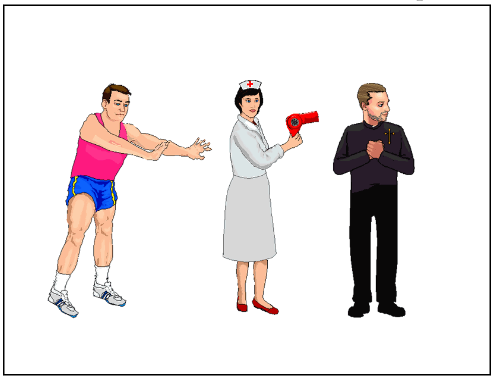
Die Amazone. . .
. . . besprüht gerade den Fußballspieler.
. . . erdolcht gerade der Mechaniker.



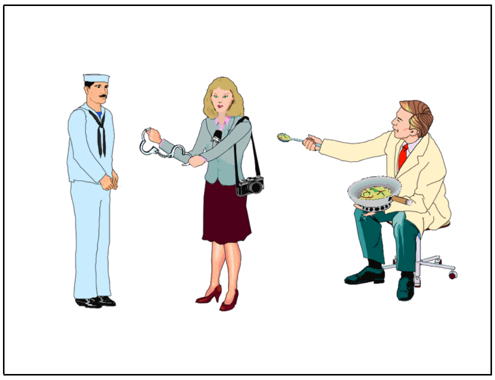
Die Krankenschwester. . .
. . . schubst in diesem Moment den Sportler.
. . . fönt in diesem Moment der Priester.



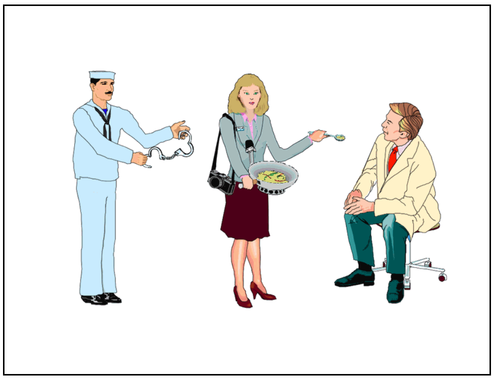
Die Krankenschwester. . .
. . . fönt in diesem Moment den Priester.
. . . schubst in diesem Moment der Sportler.



Die Journalistin. . .
. . . fesselt in diesem Moment den Matrosen.
. . . füttert in diesem Moment der Oberarzt.



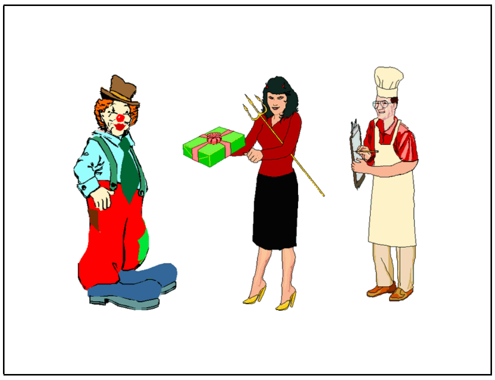
Die Journalistin. . .
. . . füttert in diesem Moment den Oberarzt.
. . . fesselt in diesem Moment der Matrose.



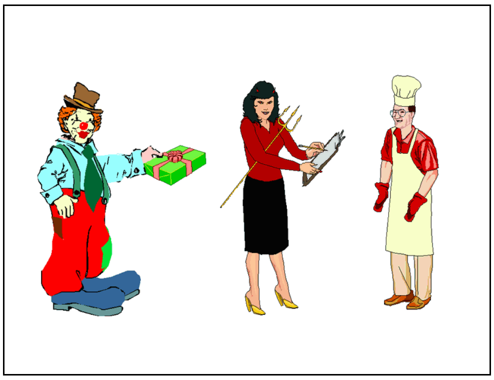
Die Bauarbeiterin...
...attackiert offensichtlich den Cellisten.
...interviewt offensichtlich der Golfer.

Die Bauarbeiterin...
...interviewt offensichtlich den Golfer.
...attackiert offensichtlich der Cellist.

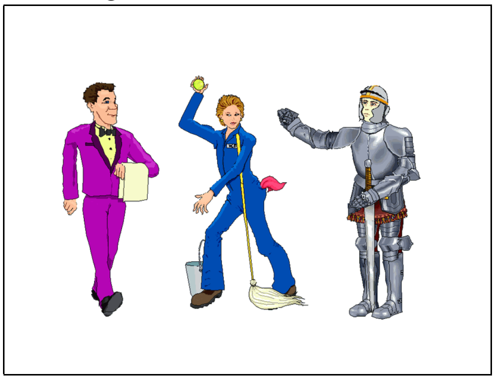
Die Teufelin...
...beschenkt in diesem Moment den Clown.
...skizziert in diesem Moment der Koch.



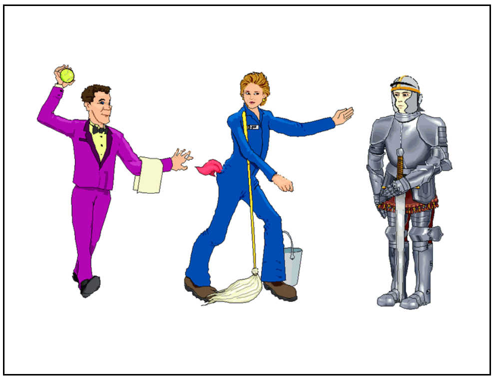
Die Teufelin...
...skizziert in diesem Moment den Koch.
...beschenkt in diesem Moment der Clown.



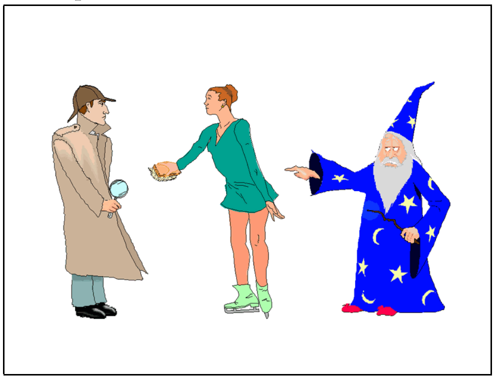
Die Putzfrau...
...bewirft soeben den Kellner.
...ohrfeigt soeben der Ritter.



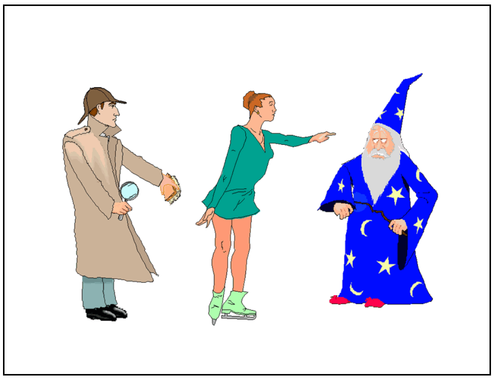
Die Putzfrau...
...ohrfeigt soeben den Ritter.
...bewirft soeben der Kellner.



Die Schlittschuhläuferin...
...schrubbt mal eben den Detektiv.
...stupst mal eben der Zauberer.



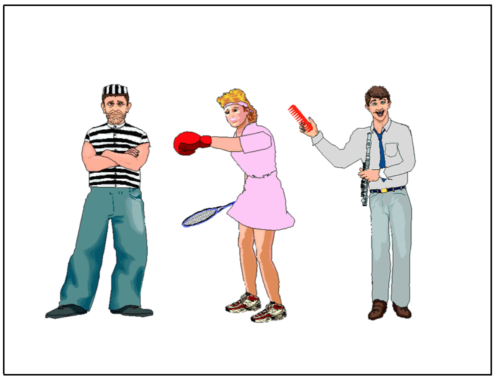
Die Schlittschuhläuferin...
...stupst mal eben den Zauberer.
...schrubbt mal eben der Detektiv.



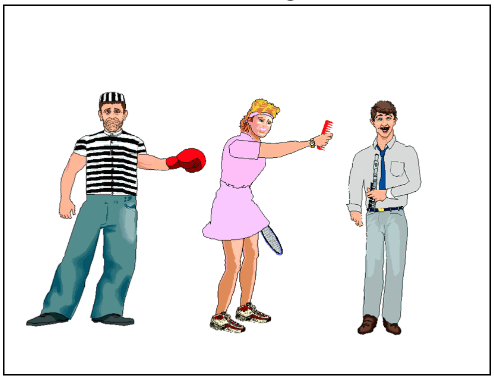
Das Dienstmädchen...
...parfümiert in diesem Moment den Henker.
...bandagiert in diesem Moment der Trommler.

Das Dienstmädchen...
...bandagiert in diesem Moment den Trommler.
...parfümiert in diesem Moment der Henker.

Die Tennisspielerin...
...boxt hier den Sträfling.
...kämmt hier der Flötist.



Die Tennisspielerin...
...kämmt hier den Flötisten.
...boxt hier der Sträfling.



Die Meerjungfrau...
...krönt gerade den Studenten.
...zupft gerade der Soldat.

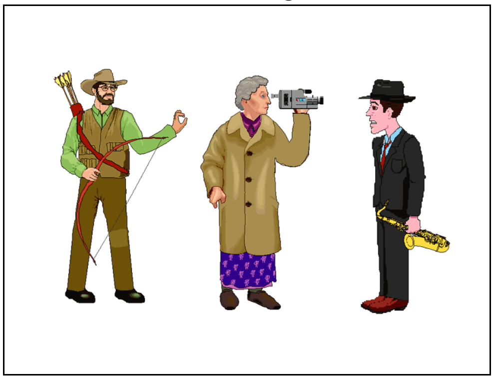
Die Meerjungfrau...
...zupft gerade den Soldaten.
...krönt gerade der Student.

Die Nonne...
...impft gerade den Schülerlotsen.
...zwickt gerade der Klarinettist.

Die Nonne...
...zwickt gerade den Klarinettisten.
...impft gerade der Schülerlotse.

Die Oma...
...kratzt soeben den Bogenschützen.
...filmt soeben der Saxophonist.

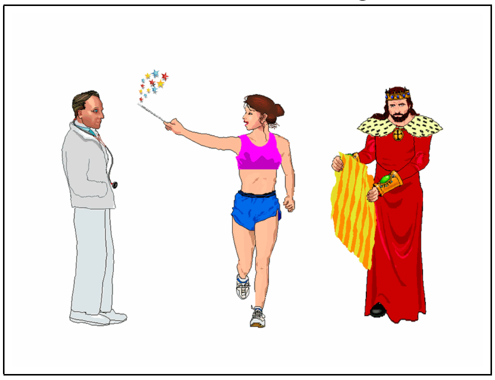
Die Oma...
...filmt soeben den Saxphonisten.
...kratzt soeben der Bogenschütze.



Die Fee. . .
. . . bürstet hier den Gangster.
. . . bespritzt hier der Tourist.

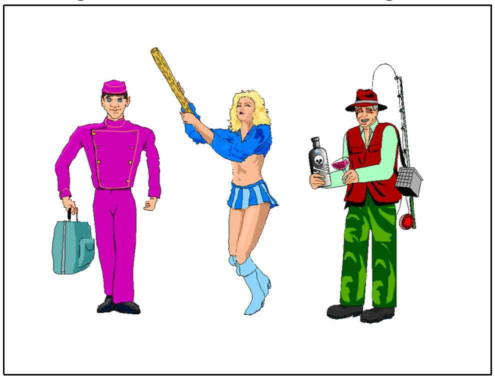
Die Fee. . .
. . . bespritzt hier den Touristen.
. . . bürstet hier der Gangster.

Die Joggerin. . .
. . . verhext mal eben den Doktor.
. . . frottiert mal eben der König.

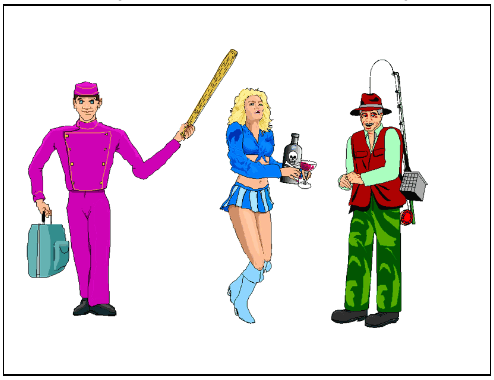


Die Joggerin. . .
. . . frottiert mal eben den König.
. . . verhext mal eben der Doktor.

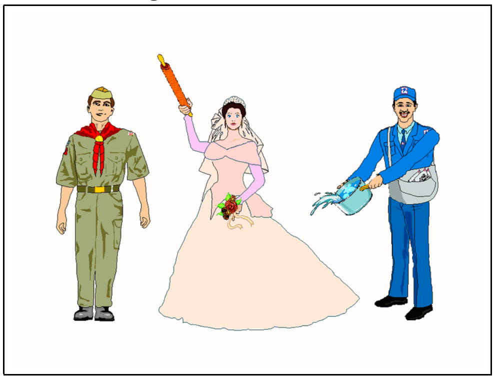
Die Cheerleaderin. . .
. . . verprügelt offensichtlich den Pagen.
. . . vergiftet offensichtlich der Angler.



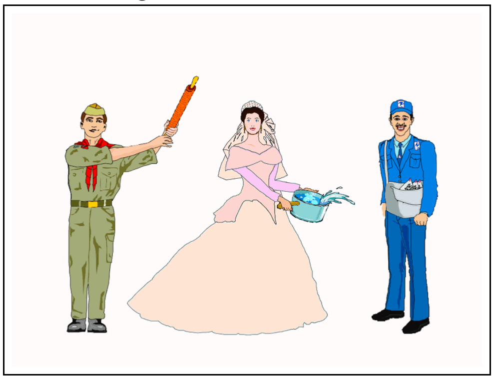
Die Cheerleaderin. . .
. . . vergiftet offensichtlich den Angler.
. . . verprügelt offensichtlich der Page.



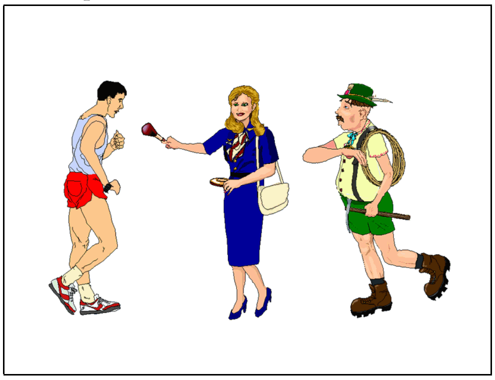
Die Braut...
...verhaut gerade den Pfadfinder.
...verbrüht gerade der Postbote.



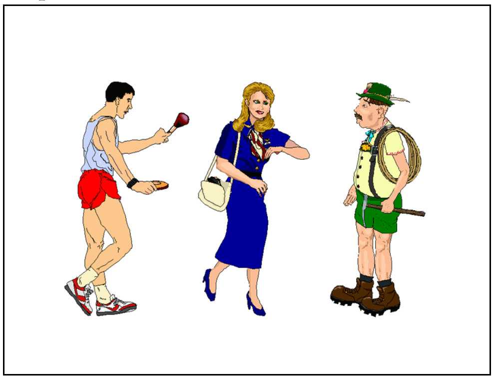
Die Braut...
...verbrüht gerade den Postboten.
...verhaut gerade der Pfadfinder.



Die Stewardess...
...pudert soeben den Leichtathleten.
...rempelt soeben der Wanderer.

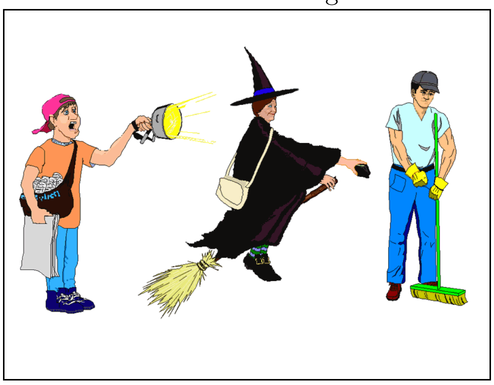


Die Stewardess...
...rempelt soeben den Wanderer.
...pudert soeben der Leichtathlet.



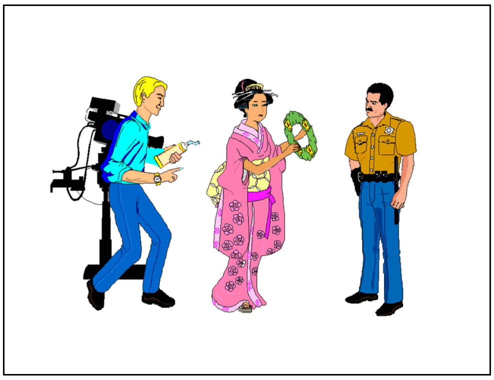
Die Hexe...
...bestrahlt hier den Zeitungsverkäufer.
...bestiehlt hier der Strassenkehrer.

Die Hexe...
...bestiehlt hier den Strassenkehrer.
...bestrahlt hier der Zeitungsverkäufer.



Die Japanerin. . .
. . . beschmiert mal eben den Kameramann.
. . . bekränzt mal eben der Ordnunghüter.

Die Japanerin. . .
. . . bekränzt mal eben den Ordnungshüter.
. . . beschmiert mal eben der Kameramann.



Die Geschäftsfrau. . .
. . . umgürtet mal eben den Klempner.
. . . verköstigt mal eben der Imker.

Die Geschäftsfrau. . .
. . . verköstigt mal eben den Imker.
. . . umgürtet mal eben der Klempner.

Die Trainerin. . .
. . . verwarnt offensichtlich den Astronauten.
. . . bekocht offensichtlich der Handwerker.

Die Trainerin. . .
. . . bekocht offensichtlich den Handwerker.
. . . verwarnt offensichtlich der Astronaut.

# Appendix B

# The data in more detail

## B.1 Fixation histograms

This section contains the full, all-measures, versions of the histograms presented in section 3.2. Those are pretty crowded, showing the percentage of Nth fixations for up to 15 regions concurrently. Region group data are in black, image regions in green (labels based on activity of the depicted characters) or blue (labels based on absolute position of depicted characters or on sentence role related to the depicted character). The N-CHAR region is shown as impulses, the background region as boxes. This is done to keep the histograms a bit more readable when printed in black and white.

Not all histograms show all values, because there is some overlap: The histograms about SVO / OVS with match / mismatch interactions, for example, do not show the S-CHAR, O-CHAR and N-CHAR values as those are identical to the agent, ambiguous char and patient regions, with assignments shown in table B.1. This table contains a full list of overlapping region names sorted by histogram condition.

Table B.1: Overlapping of region names per condition: For example for all SVO histograms, S-CHAR is not shown separately, as it refers to the same region as “ambiguous / middle character”.

| | | |
|---|---|---|
| all SVO | ambiguous char = S-CHAR | |
| all OVS | ambiguous char = O-CHAR | |
| all to-right | agent is left char | patient is right char |
| all to-left | agent is right char | patient is left char |
| SVO match | agent is N-CHAR | patient √is O-CHAR |
| SVO mismatch | agent ⋆is O-CHAR | patient is N-CHAR |
| OVS match | agent √is S-CHAR | patient is N-CHAR |
| OVS mismatch | agent is N-CHAR | patient ⋆is S-CHAR |

### B.1.1 Match / Mismatch and SVO / OVS

Figure B.1: Full histograms of fixation frequency by fixation number: Totals for all (correctly answered) SVO and OVS trials, respectively.

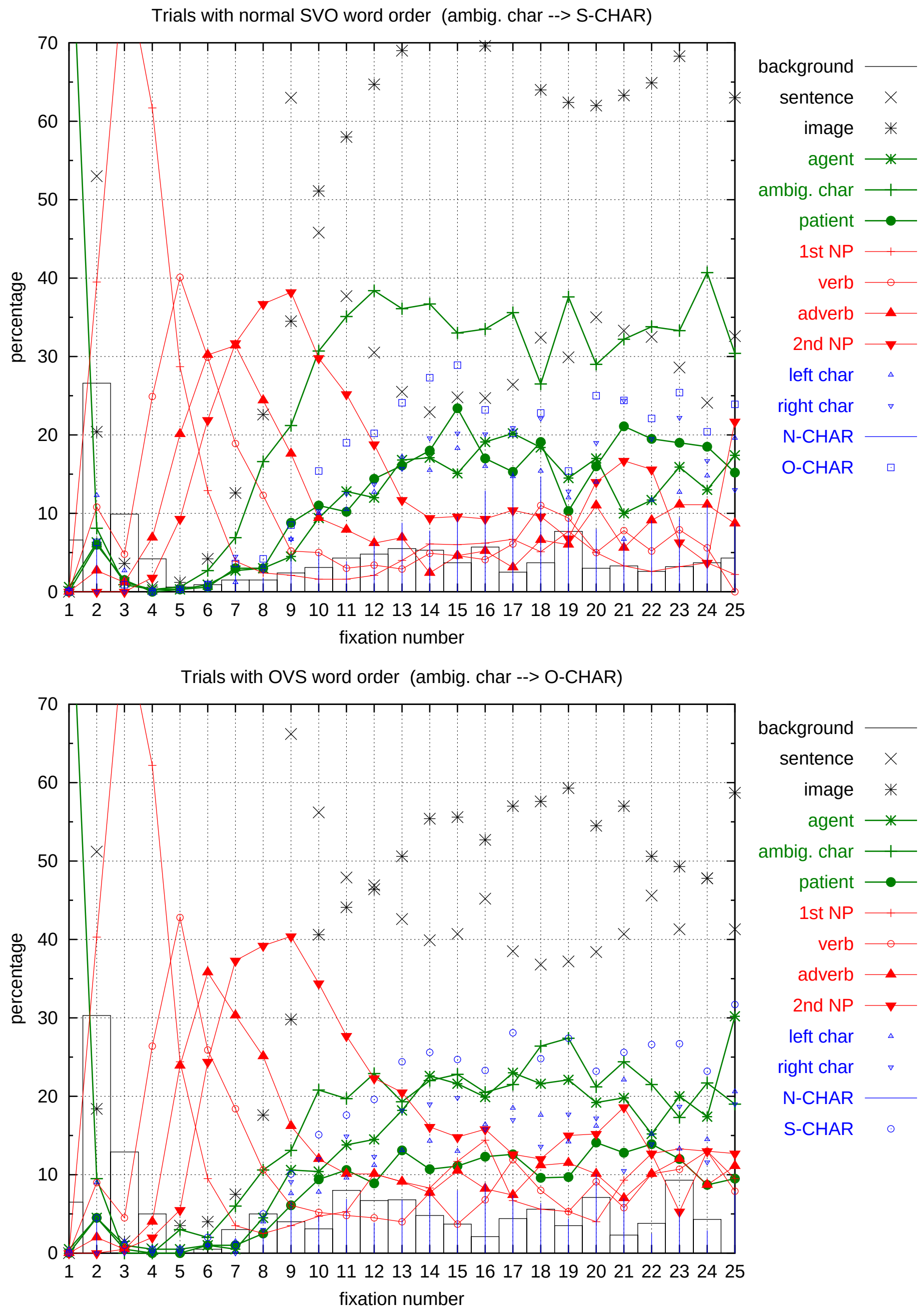
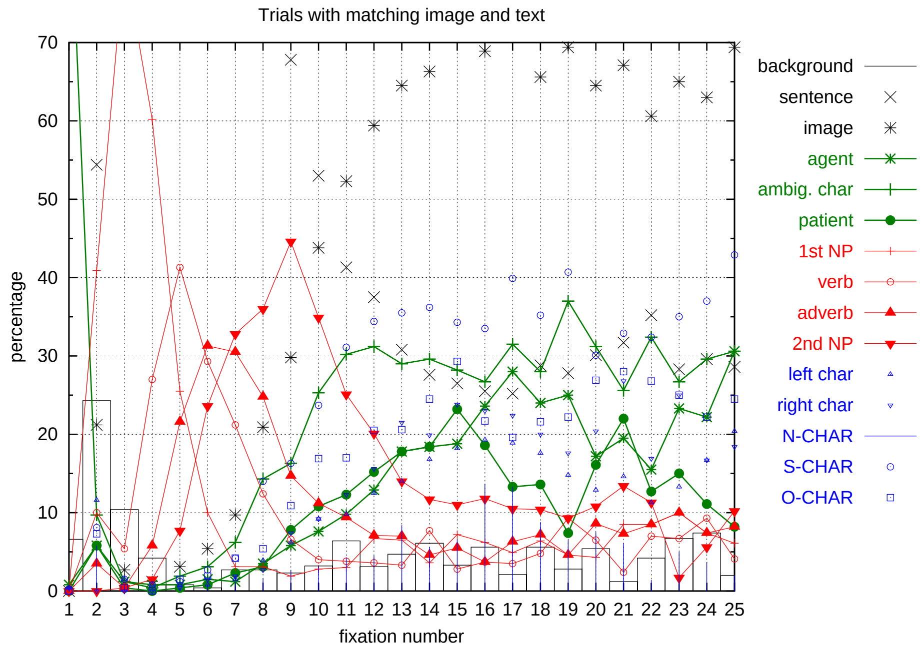
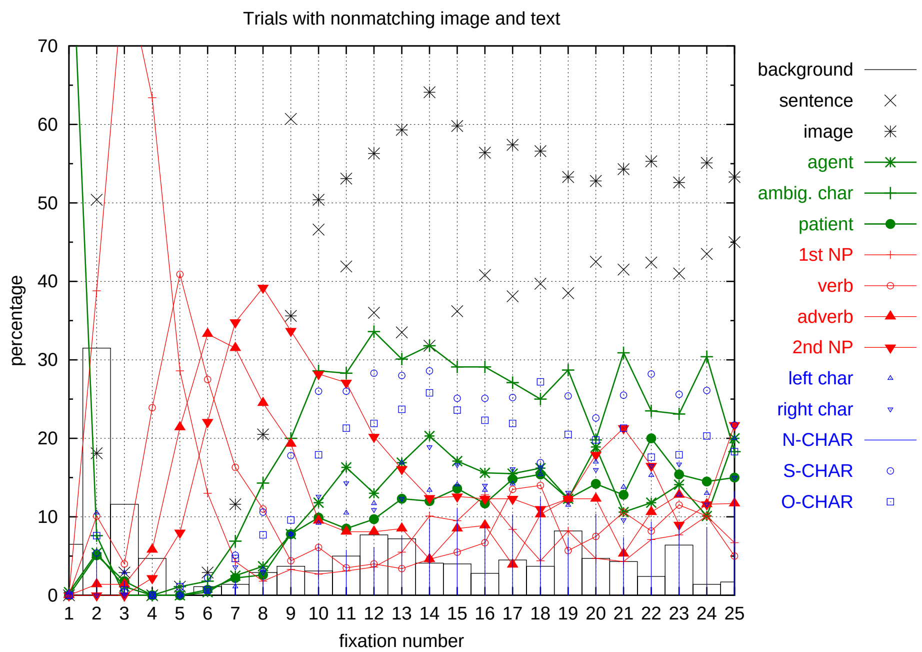


Figure B.2: Full histograms of fixation frequency by fixation number: Totals for all (correctly answered) match and mismatch trials, respectively.



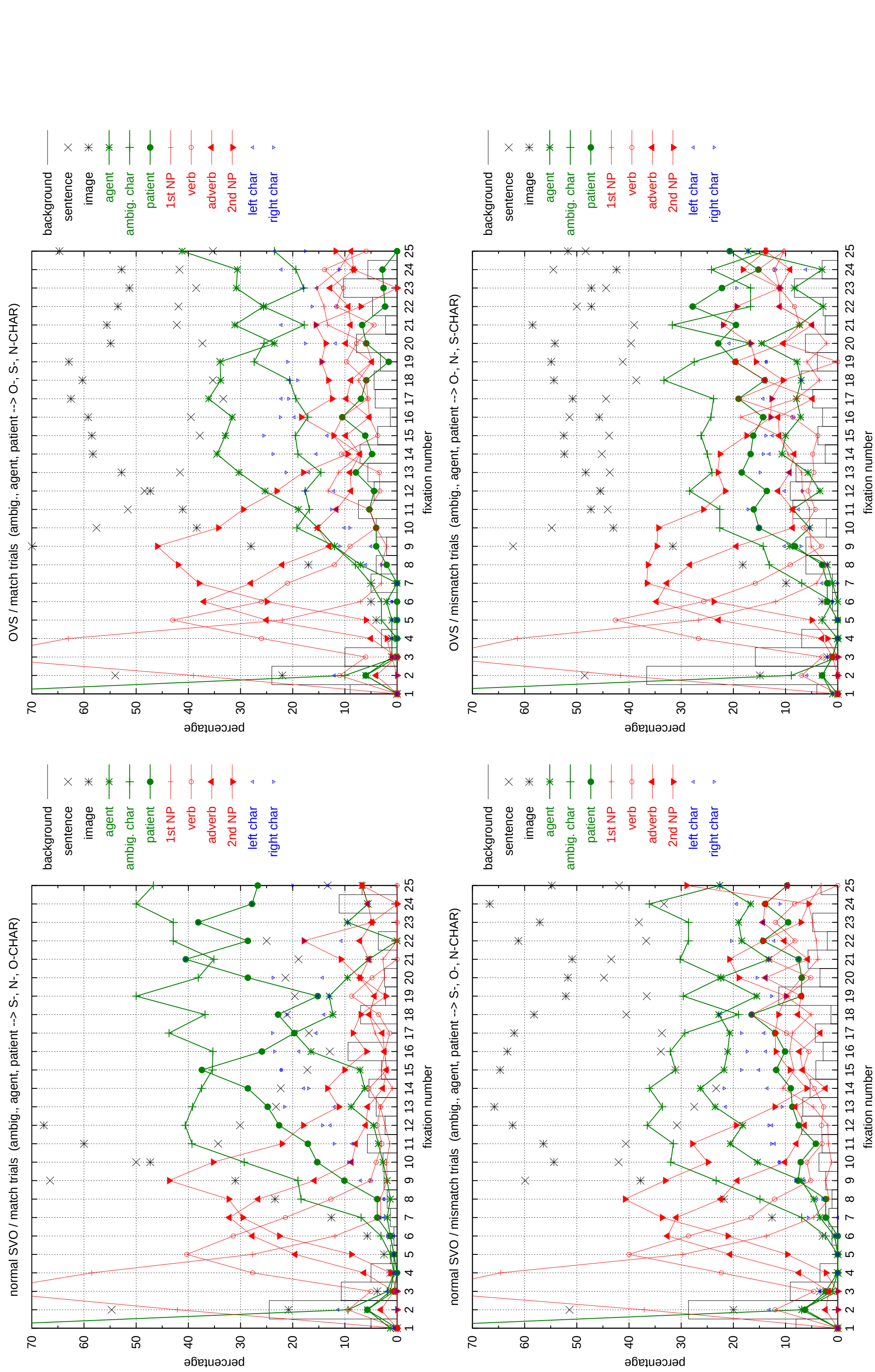
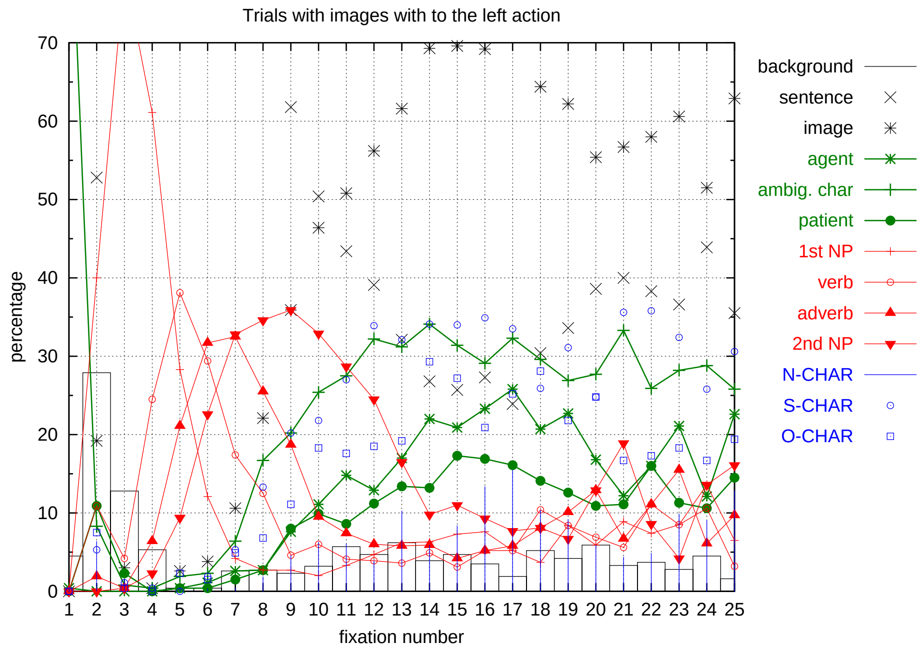
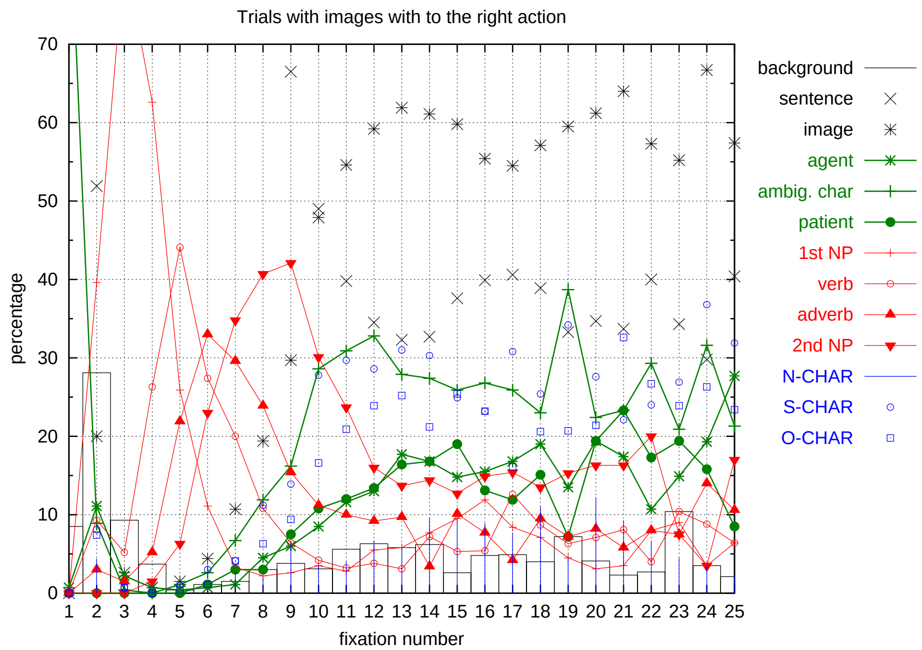



Figure B.3: Full histograms of fixation frequency by fixation number: Totals for all combinations of SVO / OVS and match / mismatch cases.

### B.1.2 Left-direction versus right-direction images

Figure B.4: Full histograms of fixation frequency by fixation number: Totals for all (correctly answered) trials with to-the-left and to-the-right image action direction, respectively.



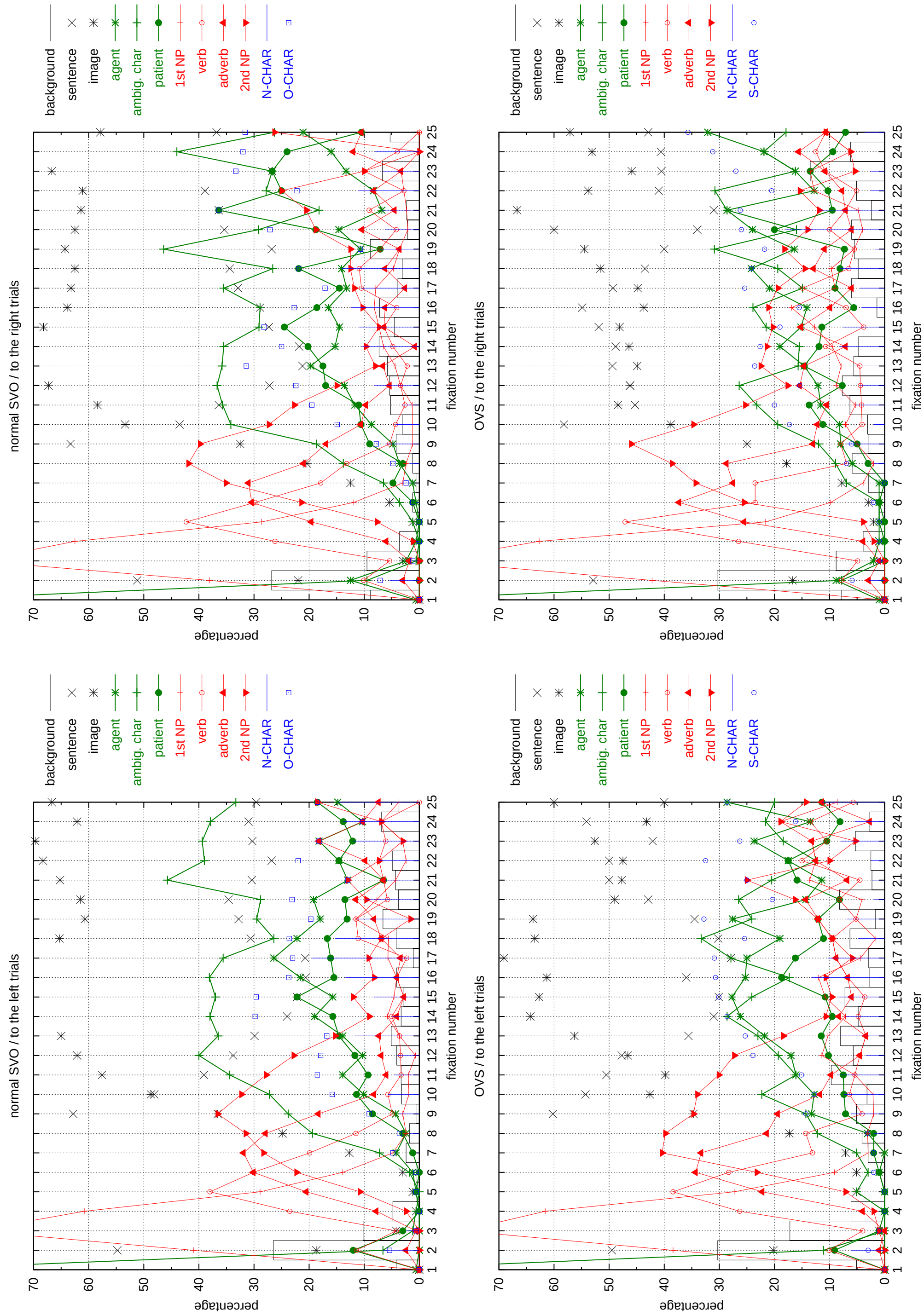



Figure B.5: Full histograms of fixation frequency by fixation number: Totals for all combinations of SVO / OVS with both image directions.

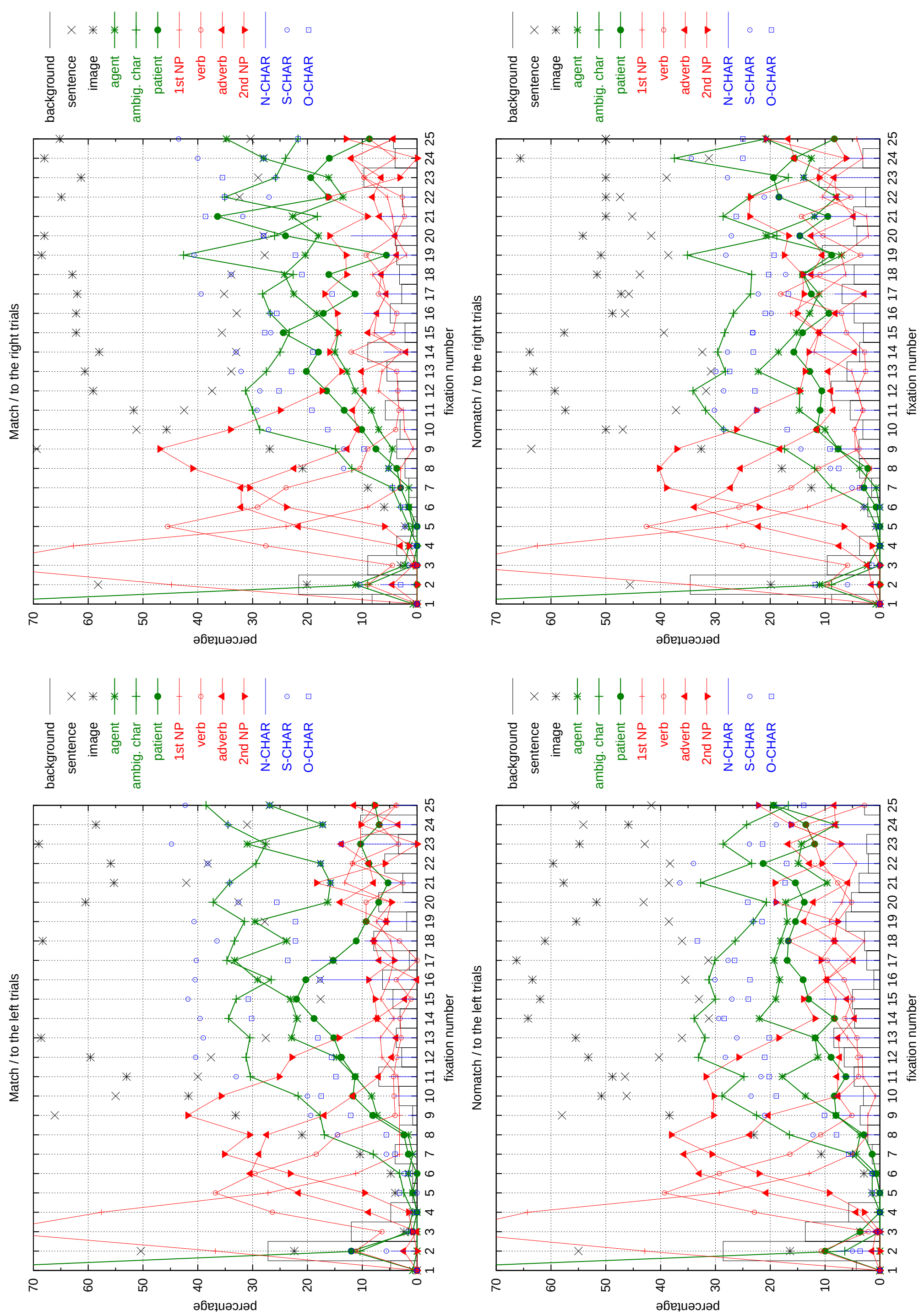



Figure B.6: Full histograms of fixation frequency by fixation number: Totals for all combinations of match / mismatch with both image directions.

## B.2 Additional effects: Image action direction in incorrectly answered trials

Please refer to section 3.2.7 for explanations on how the following two tables are structured. They give information about the effect of image action direction (to the left / to the right) specific to the *incorrectly answered* trials. Table B.3 also shows the interaction of this effect with the SVO / OVS contrast, while table B.2 adds information about the interaction with the match / mismatch contrast.

Table B.2: Significance of to-right / to-left effects and left / right asymmetries in match (√) / mismatch (⋆) effects in **incorrectly answered** trials

| **Region** / **measure** | Is ⇒≉⇐? over-all | only for √ | only for ⋆ | match ≉ mismatch? significances for ⇒ | for ⇐ | **Averages**, sorted (R/Y is ⇒/match, L/N is ⇐/mismatch) |
|---|---|---|---|---|---|---|
| **background** | | | | | | |
| time | 0.05 | — | — | — | — | L/N 133.1 ≀ L/Y 165.3 ≀ R/Y 236.4 ≀ R/N 252.1 |
| # of fix. | 0.01 | — | 0.01 | — | — | L/N 0.738 • L/Y 0.942 • R/N 1.289 ≀ R/Y 1.447 |
| **patient** | | | | | | |
| time | — | — | — | — | 0.01 | L/Y 364.6 ≀ R/Y 369.7 • R/N 606.8 • L/N 838.5 |
| # of fix. | — | — | — | — | 0.01 | L/Y 1.288 • R/Y 1.362 • R/N 2.333 • L/N 2.952 |
| 1st pass # | — | — | — | 0.01 | — | R/Y 0.915 ≀ L/Y 0.942 • L/N 1.238 ≀ R/N 1.333 |
| **S-CHAR** | | | | | | |
| 1st fix. time | — | — | — | — | — | R/Y 175.5 ≀ L/Y 207.3 • R/N 259.4 ≀ L/N 270.4 |
| **N-CHAR** | | | | | | |
| 1st fix. time | — | — | — | 0.01 | — | R/N 115.3 ≀ L/N 131.5 • L/Y 150.2 ≀ R/Y 220.5 |
| **left pic.** | | | | | | |
| time | — | — | 0.01 | — | 0.01 | R/N 335.8 ≀ L/Y 364.6 • R/Y 513.8 • L/N 838.5 |
| # of fix. | — | — | 0.01 | — | 0.01 | L/Y 1.288 ≀ R/N 1.400 • R/Y 1.894 • L/N 2.952 |
| 1st fix. time | 0.01 | — | 0.001 | — | — | R/N 120.6 ≀ R/Y 167.0 • L/Y 184.2 ≀ L/N 244.1 |
| **right pic.** | | | | | | |
| time | — | — | 0.01 | — | — | L/N 291.3 ≀ R/Y 369.7 • L/Y 399.4 ≀ R/N 606.8 |
| # of fix. | — | — | 0.01 | — | — | L/N 1.095 ≀ R/Y 1.362 • L/Y 1.577 ≀ R/N 2.333 |
| 1st fix. time | 0.01 | — | 0.01 | — | — | L/N 141.2 ≀ L/Y 158.7 • R/Y 207.0 ≀ R/N 255.4 |
| 1st pass # | — | — | — | 0.01 | — | L/N 0.905 ≀ R/Y 0.915 ≀ L/Y 1.173 • R/N 1.333 |

fix.: fixation(s), #: number (of fixations), ins.: inspection, r-path: regression path

Underlined: significant effect in “nonstandard” direction (OVS ≺ SVO or ⇒≺⇐)

Table B.3: Significance of to-right / to-left effects and left / right asymmetries in SVO / OVS effects in **incorrectly answered** trials

| **Region** / **measure** | Is ⇒≉⇐? over-all | only for SVO | only for OVS | SVO ≉ OVS? significances for ⇒ | for ⇐ | **Averages**, sorted (R/S is ⇒/SVO, L/O is ⇐/OVS) |
|---|---|---|---|---|---|---|
| **background** | | | | | | |
| time | 0.05 | — | — | — | — | L/O 141.4 ≀ L/S 197.4 • R/O 226.1 ≀ R/S 336.1 |
| # of fix. | 0.01 | 0.01 | — | — | — | L/O 0.846 ≀ L/S 0.875 • R/O 1.247 • R/S 2.000 |
| 1st fix. time | — | — | — | — | 0.01 | L/O 75.49 ≀ R/O 108.4 • R/S 132.9 ≀ L/S 160.6 |
| **verb** | | | | | | |
| 1st fix. time | — | — | — | — | — | R/S 156.4 ≀ L/S 176.8 • R/O 213.2 ≀ L/O 224.0 |
| **adverb** | | | | | | |
| 1st fix. time | — | — | — | — | — | R/S 149.1 ≀ R/O 214.9 • L/S 228.7 ≀ L/O 230.0 |
| **agent** | | | | | | |
| 1st fix. time | — | — | — | 0.01 | — | L/S 122.2 ≀ R/O 129.6 ≀ L/O 156.8 • R/S 219.6 |
| **S-CHAR** | | | | | | |
| time | — | — | — | — | 0.01 | R/O 598.2 ≀ L/O 684.0 • R/S 929.0 • L/S 1295 |
| 1st pass time | — | 0.01 | — | — | 0.001 | R/O 330.7 ≀ R/S 332.9 ∘ L/O 345.8 ≺ L/S 737.4 |
| 1st pass # | — | — | — | — | — | R/S 1.200 ≀ R/O 1.273 ≀ L/O 1.321 • L/S 2.125 |
| **left pic.** | | | | | | |
| 1st fix. time | 0.01 | — | 0.001 | 0.01 | — | R/O 129.6 • L/S 207.2 • L/O 211.8 ≀ R/S 219.6 |
| **middle pic.** | | | | | | |
| 1st pass time | — | 0.01 | — | — | — | R/S 332.9 ≀ R/O 453.1 • L/O 470.3 • L/S 737.4 |
| **right pic.** | | | | | | |
| 1st fix. time | 0.01 | — | 0.01 | — | — | L/S 122.2 ≀ L/O 156.8 ≀ R/S 161.6 • R/O 244.1 |

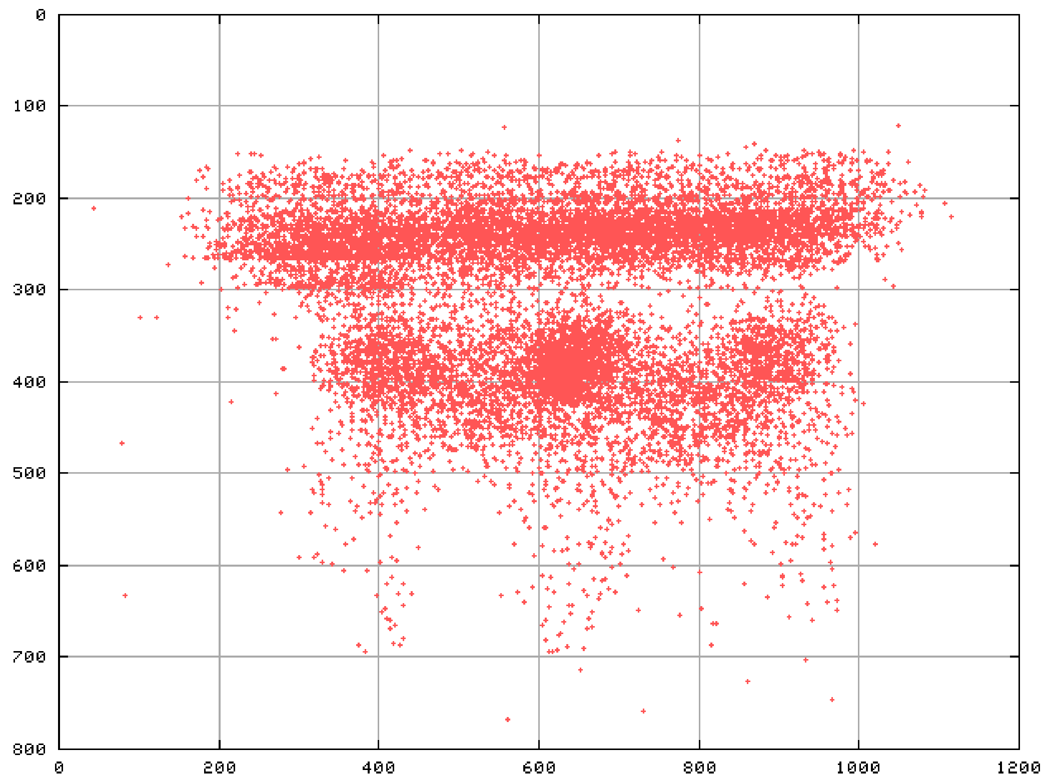
fix.: fixation(s), #: number (of fixations), ins.: inspection, r-path: regression path
Underlined: significant effect in "nonstandard" direction (OVS ≺ SVO or ⇒≺⇐)

## B.3 Scatterplot of all fixations

To answer the eternal question where people *actually* looked at during the whole experiment, this section features a scatterplot of all fixations ever recorded for our experiment. Those include the fixations during trials which had excessive track loss, and both the fixations for correctly answered and for incorrectly answered trials. However, fixations for filler items were not recorded at all in the first place.

Figure B.7 shows, on a coordinate system which has the same size as the screen canvas visible to the participants, a scatterplot of all recorded fixations: You can see that the bulk of fixations went to the text regions and to the heads, hands and tools of the depicted characters. The heavier blob at the center is a result of us forcing the participants to look at the center of the screen to start each trial. The vertical location of the text varied somewhat depending on how high the depicted scenes on each of the 48 stimulus pictures were, as the text location is selected relative to the image contents. The filler items had more variation in image height, so the participants had to look at the overall display contents and find the exact text location each time. Similarily, the text was centered horizontally, but text length varied.

Figure B.7: Overall scatterplot of fixations: This scatterplot shows all captured trial (not filler) fixations during the whole experiment, both of correctly answered and incorrectly answered trials, even for trials with excessive track loss. Note that the position of the depicted characters varied to some degree among the 48 stimulus pictures, which also caused some variations in the vertical placement of the corresponding stimulus text.



## B.4 List of all significant effects

The automatically generated list of all significant effects, their significance levels, t-values of the t-tests, means, standard deviations and numbers of items of all tested conditions: Values are listed as "$condition = mean\ (items,\ stddev)$". The results of t-tests are listed as either "$t = tValue\ (p < value)$" or just as "$conditions = tValue$". In the latter case, values are only listed if $p < 0.01$ (the corrected threshold for our EDA), and are underlined if $p < 0.001$. The *. and ** prefixes mark contrasts which contain $p < 0.01$ and $p < 0.001$ effects, respectively.

TWO is an effect along one of the two main contrasts – SVO versus OVS, match versus mismatch – or an effect of the balancing factor of image action direction to the left versus to the right.

FOUR is a list of effects for a combination of two of the main contrasts or of one of the main contrasts and the image action direction: Four means are compared by doing all six t-tests for all combinations of two of the four values. The list shows the values sorted from small to big, using the same symbols as the pretty-printed effect tables, explained in table 3.6. The mean / number / standard deviation triples are wrapped into two lines in the FOUR list, and we use abbreviations of categories: For example "sn" is SVO / mismatch and "snsy" is about the SVO mismatch to SVO match t-test results. Used abbreviation letters are: S, O, N, Y, R, L, for SVO, OVS, mismatch, match, to-the-right and to-the-left, respectively.

### B.4.1 Effects in correctly answered trials

** TWO, $t = 5.886$ ($p < 0.001$) button time:
$SVO = 4898.150\ (334,\ 1917.120) \leftarrow OVS = 6125.035\ (201,\ 2900.304)$
* FOUR button time: $\substack{sy=4605.233\\(159,\ 1452.901)} \prec \substack{sn=5164.286\\(175,\ 2228.768)} \circ \substack{on=5928.198\\(101,\ 2675.318)} \bullet \substack{oy=6323.840\\(100,\ 3111.974)}$
$sysn = 2.686$ $\underline{syon = 5.156}$ $\underline{syoy = 6.005}$ $\underline{snoy = 3.579}$ 2nd<4th
* FOUR button time: $\substack{rs=4842.833\\(168,\ 1746.461)} \circ \substack{ls=4954.133\\(166,\ 2079.461)} \prec \substack{ro=6111.784\\(102,\ 2985.238)} \circ \substack{lo=6138.687\\(99,\ 2825.250)}$
$\underline{rsro = 4.408}$ $\underline{rslo = 4.633}$ $\underline{lsro = 3.736}$ $\underline{lslo = 3.912}$

** TWO, $t = 5.593$ ($p < 0.001$) anywhere time:
$SVO = 4196.383\ (334,\ 1759.632) \leftarrow OVS = 5270.836\ (201,\ 2680.677)$
* FOUR anywhere time: $\substack{sy=3920.748\\(159,\ 1328.121)} \prec \substack{sn=4446.817\\(175,\ 2047.309)} \circ \substack{on=5103.881\\(101,\ 2451.906)} \bullet \substack{oy=5439.460\\(100,\ 2896.119)}$
$sysn = 2.756$ $\underline{syon = 5.035}$ $\underline{syoy = 5.728}$ $snoy = 3.313$ 2nd<4th
* FOUR anywhere time: $\substack{rs=4165.661\\(168,\ 1590.806)} \circ \substack{ls=4227.476\\(166,\ 1919.800)} \prec \substack{ro=5263.284\\(102,\ 2770.212)} \circ \substack{lo=5278.616\\(99,\ 2599.278)}$
$\underline{rsro = 4.136}$ $\underline{rslo = 4.342}$ $\underline{lsro = 3.610}$ $\underline{lslo = 3.767}$

** TWO, $t = 5.132$ ($p < 0.001$) anywhere # of fix.:
$SVO = 18.222\ (334,\ 7.985) \leftarrow OVS = 22.552\ (201,\ 11.488)$
* FOUR anywhere # of fix.: $\substack{sy=16.994\\(159,\ 6.234)} \prec \substack{sn=19.337\\(175,\ 9.171)} \circ \substack{on=21.891\\(101,\ 11.425)} \bullet \substack{oy=23.220\\(100,\ 11.570)}$
$sysn = 2.704$ $\underline{syon = 4.462}$ $\underline{syoy = 5.616}$ $snoy = 3.065$ 2nd<4th
* FOUR anywhere # of fix.: $\substack{rs=17.827\\(168,\ 7.363)} \circ \substack{ls=18.620\\(166,\ 8.573)} \prec \substack{lo=22.525\\(99,\ 10.391)} \circ \substack{ro=22.578\\(102,\ 12.513)}$

$\underline{rslo = 4.308}$ $\underline{rsro = 3.929}$ $lslo = 3.309$ $lsro = 3.070$

*. TWO, $t = 3.221$ ($p < 0.01$) background time:
$SVO = 169.734\ (334,\ 217.852) \leftarrow OVS = 246.264\ (201,\ 331.372)$
. FOUR background time: $\underset{(159,\ 195.262)}{sy{=}157.409} \wr \underset{(175,\ 236.520)}{sn{=}180.931} \bullet \underset{(100,\ 362.998)}{oy{=}241.850} \wr \underset{(101,\ 298.540)}{on{=}250.634}$
$syon = 3.045$ 1st<4th
. FOUR background time: $\underset{(166,\ 177.146)}{ls{=}152.373} \bullet \underset{(168,\ 251.067)}{rs{=}186.887} \bullet \underset{(102,\ 336.190)}{ro{=}233.873} \wr \underset{(99,\ 327.550)}{lo{=}259.030}$
$lsro = 2.594$ $lslo = 3.439$ 1st<3rd-4th

** TWO, $t = 3.729$ ($p < 0.001$) background # of fix.:
$SVO = 0.934\ (334,\ 0.959) \leftarrow OVS = 1.338\ (201,\ 1.548)$
* FOUR background # of fix.: $\underset{(159,\ 0.918)}{sy{=}0.843} \bullet \underset{(175,\ 0.991)}{sn{=}1.017} \bullet \underset{(100,\ 1.709)}{oy{=}1.300} \wr \underset{(101,\ 1.377)}{on{=}1.376}$
$syoy = 2.796$ $\underline{syon = 3.748}$ 1st<3rd-4th
* FOUR background # of fix.: $\underset{(166,\ 0.879)}{ls{=}0.880} \wr \underset{(168,\ 1.032)}{rs{=}0.988} \bullet \underset{(102,\ 1.551)}{ro{=}1.245} \bullet \underset{(99,\ 1.546)}{lo{=}1.434}$
$\underline{lslo = 3.725}$ $rslo = 2.824$ 1st-2nd<4th

.. TWO, $t = 2.535$ ($p < 0.05$) image time:
$SVO = 1999.805\ (334,\ 1247.336) \leftarrow OVS = 2313.766\ (201,\ 1593.108)$
* FOUR image time: $\underset{(159,\ 954.601)}{sy{=}1905.887} \wr \underset{(175,\ 1461.067)}{sn{=}2085.137} \bullet \underset{(101,\ 1380.526)}{on{=}2127.099} \wr \underset{(100,\ 1769.379)}{oy{=}2502.300}$
$\underline{syoy = 3.516}$ 1st<4th
. FOUR image time: $\underset{(168,\ 1001.160)}{rs{=}1945.839} \wr \underset{(166,\ 1455.865)}{ls{=}2054.422} \bullet \underset{(102,\ 1520.063)}{ro{=}2153.147} \wr \underset{(99,\ 1656.423)}{lo{=}2479.253}$
$rslo = 3.281$ 1st<4th

.. TWO, $t = 2.146$ ($p < 0.05$) image # of fix.:
$SVO = 7.859\ (334,\ 4.855) \leftarrow OVS = 8.881\ (201,\ 6.040)$
. FOUR image # of fix.: $\underset{(159,\ 4.103)}{sy{=}7.516} \wr \underset{(175,\ 5.443)}{sn{=}8.171} \bullet \underset{(101,\ 5.982)}{on{=}8.317} \wr \underset{(100,\ 6.076)}{oy{=}9.450}$
$syoy = 3.057$ 1st<4th
. FOUR image # of fix.: $\underset{(168,\ 3.982)}{rs{=}7.512} \wr \underset{(166,\ 5.593)}{ls{=}8.211} \bullet \underset{(102,\ 6.080)}{ro{=}8.363} \wr \underset{(99,\ 5.983)}{lo{=}9.414}$
$rslo = 3.115$ 1st<4th

** TWO, $t = 6.271$ ($p < 0.001$) sentence time:
$SVO = 2026.844\ (334,\ 914.309) \leftarrow OVS = 2710.806\ (201,\ 1608.027)$
* FOUR sentence time: $\underset{(159,\ 733.018)}{sy{=}1857.453} \prec \underset{(175,\ 1030.489)}{sn{=}2180.749} \prec \underset{(100,\ 1634.952)}{oy{=}2695.310} \circ \underset{(101,\ 1588.927)}{on{=}2726.149}$
$sysn = 3.274$ $\underline{syoy = 5.629}$ $\underline{syon = 5.970}$ $snoy = 3.199$ $snon = 3.455$
* FOUR sentence time: $\underset{(166,\ 859.192)}{ls{=}2020.681} \circ \underset{(168,\ 968.231)}{rs{=}2032.935} \prec \underset{(99,\ 1504.568)}{lo{=}2540.333} \circ \underset{(102,\ 1693.423)}{ro{=}2876.265}$
$\underline{lslo = 3.580}$ $\underline{lsro = 5.468}$ $rslo = 3.351$ $\underline{rsro = 5.207}$

** TWO, $t = 5.980$ ($p < 0.001$) sentence # of fix.:
$SVO = 9.428\ (334,\ 4.319) \leftarrow OVS = 12.333\ (201,\ 6.918)$
* FOUR sentence # of fix.: $\underset{(159,\ 3.221)}{sy{=}8.635} \prec \underset{(175,\ 5.018)}{sn{=}10.149} \prec \underset{(101,\ 6.943)}{on{=}12.198} \circ \underset{(100,\ 6.926)}{oy{=}12.470}$
$sysn = 3.244$ $\underline{syon = 5.596}$ $\underline{syoy = 6.027}$ $snon = 2.830$ $snoy = 3.202$
* FOUR sentence # of fix.: $\underset{(168,\ 4.340)}{rs{=}9.327} \circ \underset{(166,\ 4.309)}{ls{=}9.530} \prec \underset{(99,\ 5.972)}{lo{=}11.677} \circ \underset{(102,\ 7.703)}{ro{=}12.971}$

$\underline{rslo = 3.704}$ $\underline{rsro = 4.971}$ $lslo = 3.385$ $\underline{lsro = 4.687}$

** TWO, $t = 5.259$ ($p < 0.001$) agent time:
$SVO = 413.266\ (334,\ 535.090) \leftarrow OVS = 741.736\ (201,\ 909.842)$
* FOUR agent time: $\underset{(159,\ 266.335)}{sy{=}182.566} \circ \underset{(101,\ 297.337)}{on{=}264.416} \prec \underset{(175,\ 625.081)}{sn{=}622.874} \prec \underset{(100,\ 1056.337)}{oy{=}1223.830}$
$\underline{sysn = 8.229}$ $\underline{syoy = 11.857}$ $\underline{onsn = 5.417}$ $\underline{onoy = 8.783}$ $\underline{snoy = 5.929}$
* FOUR agent time: $\underset{(168,\ 474.853)}{rs{=}403.619} \circ \underset{(166,\ 591.134)}{ls{=}423.030} \prec \underset{(102,\ 871.727)}{ro{=}727.324} \circ \underset{(99,\ 951.738)}{lo{=}756.586}$
$\underline{rsro = 3.947}$ $\underline{rslo = 4.033}$ $lsro = 3.403$ $\underline{lslo = 3.520}$

** TWO, $t = 5.180$ ($p < 0.001$) agent # of fix.:
$SVO = 1.617\ (334,\ 2.032) \leftarrow OVS = 2.726\ (201,\ 2.910)$
* FOUR agent # of fix.: $\underset{(159,\ 1.008)}{sy{=}0.767} \prec \underset{(101,\ 1.291)}{on{=}1.149} \prec \underset{(175,\ 2.392)}{sn{=}2.389} \prec \underset{(100,\ 3.213)}{oy{=}4.320}$
$syon = 2.661$ $\underline{sysn = 7.930}$ $\underline{syoy = 12.979}$ $\underline{onsn = 4.818}$ $\underline{onoy = 9.199}$ $\underline{snoy = 5.668}$
* FOUR agent # of fix.: $\underset{(168,\ 1.751)}{rs{=}1.524} \circ \underset{(166,\ 2.283)}{ls{=}1.711} \prec \underset{(102,\ 2.615)}{ro{=}2.578} \circ \underset{(99,\ 3.192)}{lo{=}2.879}$
$\underline{rsro = 3.966}$ $\underline{rslo = 4.479}$ $lsro = 2.856$ $lslo = 3.460$

.. TWO, $t = 2.506$ ($p < 0.05$) agent 1st fix. time:
$SVO = 158.293\ (334,\ 150.564) \leftarrow OVS = 192.682\ (201,\ 158.832)$
* FOUR agent 1st fix. time: $\underset{(159,\ 142.224)}{sy{=}108.673} \circ \underset{(101,\ 132.737)}{on{=}131.406} \prec \underset{(175,\ 143.955)}{sn{=}203.377} \prec \underset{(100,\ 159.589)}{oy{=}254.570}$
$\underline{sysn = 6.039}$ $\underline{syoy = 7.664}$ $\underline{onsn = 4.115}$ $\underline{onoy = 5.951}$ $snoy = 2.726$
. FOUR agent 1st fix. time: $\underset{(166,\ 137.775)}{ls{=}157.855} \wr \underset{(168,\ 162.630)}{rs{=}158.726} \bullet \underset{(99,\ 123.965)}{lo{=}173.101} \wr \underset{(102,\ 185.214)}{ro{=}211.686}$
$lsro = 2.717$ 1st<4th

** TWO, $t = 4.790$ ($p < 0.001$) agent 1st pass time:
$SVO = 263.099\ (334,\ 289.752) \leftarrow OVS = 434.567\ (201,\ 537.321)$
* FOUR agent 1st pass time: $\underset{(159,\ 204.933)}{sy{=}146.119} \circ \underset{(101,\ 224.416)}{on{=}204.149} \prec \underset{(175,\ 314.122)}{sn{=}369.383} \prec \underset{(100,\ 650.781)}{oy{=}667.290}$
$\underline{sysn = 7.610}$ $\underline{syoy = 9.394}$ $\underline{onsn = 4.645}$ $\underline{onoy = 6.758}$ $\underline{snoy = 5.108}$
* FOUR agent 1st pass time: $\underset{(166,\ 270.553)}{ls{=}262.663} \circ \underset{(168,\ 308.361)}{rs{=}263.530} \prec \underset{(99,\ 490.143)}{lo{=}422.545} \circ \underset{(102,\ 581.658)}{ro{=}446.235}$
$lslo = 3.421$ $\underline{lsro = 3.499}$ $rslo = 3.254$ $rsro = 3.368$

** TWO, $t = 5.087$ ($p < 0.001$) agent 1st pass #:
$SVO = 1.048\ (334,\ 0.961) \leftarrow OVS = 1.617\ (201,\ 1.627)$
* FOUR agent 1st pass #: $\underset{(159,\ 0.743)}{sy{=}0.629} \circ \underset{(101,\ 0.909)}{on{=}0.881} \prec \underset{(175,\ 0.979)}{sn{=}1.429} \prec \underset{(100,\ 1.845)}{oy{=}2.360}$
$\underline{sysn = 8.345}$ $\underline{syoy = 10.558}$ $\underline{onsn = 4.591}$ $\underline{onoy = 7.219}$ $\underline{snoy = 5.469}$
* FOUR agent 1st pass #: $\underset{(168,\ 0.960)}{rs{=}1.012} \circ \underset{(166,\ 0.962)}{ls{=}1.084} \prec \underset{(102,\ 1.632)}{ro{=}1.578} \circ \underset{(99,\ 1.630)}{lo{=}1.657}$
$\underline{rsro = 3.593}$ $\underline{rslo = 4.069}$ $lsro = 3.119$ $\underline{lslo = 3.595}$

** FOUR patient time: $\underset{(100,\ 278.946)}{oy{=}205.160} \circ \underset{(175,\ 343.339)}{sn{=}287.360} \prec \underset{(159,\ 489.935)}{sy{=}575.786} \circ \underset{(101,\ 632.761)}{on{=}693.980}$
$\underline{oysy = 6.892}$ $\underline{oyon = 7.074}$ $\underline{snsy = 6.275}$ $\underline{snon = 6.922}$

** FOUR patient # of fix.: $\underset{(100,\ 1.086)}{oy{=}0.850} \circ \underset{(175,\ 1.440)}{sn{=}1.171} \prec \underset{(159,\ 1.909)}{sy{=}2.226} \circ \underset{(101,\ 2.273)}{on{=}2.436}$
$\underline{oysy = 6.569}$ $\underline{oyon = 6.299}$ $\underline{snsy = 5.733}$ $\underline{snon = 5.653}$

** FOUR patient 1st fix. time: $\underset{(100,\ 141.816)}{oy{=}119.370} \circ \underset{(175,\ 154.817)}{sn{=}158.366} \prec \underset{(159,\ 145.678)}{sy{=}227.604} \circ \underset{(101,\ 182.707)}{on{=}253.861}$
$\underline{oysy = 5.881}$ $\underline{oyon = 5.826}$ $\underline{snsy = 4.198}$ $\underline{snon = 4.616}$

** FOUR patient 1st pass time: $\underset{(100,\ 189.246)}{oy{=}150.390} \circ \underset{(175,\ 202.554)}{sn{=}189.937} \prec \underset{(159,\ 224.444)}{sy{=}324.340} \circ \underset{(101,\ 330.135)}{on{=}372.257}$
$\underline{oysy = 6.442}$ $\underline{oyon = 5.838}$ $\underline{snsy = 5.753}$ $\underline{snon = 5.686}$

** FOUR patient 1st pass #: $\underset{(100,\ 0.770)}{oy{=}0.650} \circ \underset{(175,\ 0.720)}{sn{=}0.777} \prec \underset{(159,\ 0.787)}{sy{=}1.277} \circ \underset{(101,\ 0.956)}{on{=}1.307}$
$\underline{oysy = 6.291}$ $\underline{oyon = 5.359}$ $\underline{snsy = 6.057}$ $\underline{snon = 5.205}$

** TWO, $t = 3.950$ ($p < 0.001$) left pic. time:
$SVO = 416.757\ (334,\ 480.676) \leftarrow OVS = 629.741\ (201,\ 766.602)$
* FOUR left pic. time: $\underset{(159,\ 480.989)}{sy{=}371.440} \wr \underset{(175,\ 478.040)}{sn{=}457.931} \bullet \underset{(101,\ 618.321)}{on{=}524.337} \bullet \underset{(100,\ 882.192)}{oy{=}736.200}$
$\underline{syoy = 4.299}$ $snoy = 3.394$ 1st-2nd<4th
* FOUR left pic. time: $\underset{(168,\ 474.853)}{rs{=}403.619} \wr \underset{(166,\ 487.573)}{ls{=}430.054} \bullet \underset{(99,\ 629.325)}{lo{=}529.202} \bullet \underset{(102,\ 871.727)}{ro{=}727.324}$
$\underline{rsro = 3.947}$ $\underline{lsro = 3.578}$ 1st-2nd<4th

*. TWO, $t = 3.212$ ($p < 0.01$) left pic. # of fix.:
$SVO = 1.599\ (334,\ 1.786) \leftarrow OVS = 2.184\ (201,\ 2.406)$
* FOUR left pic. # of fix.: $\underset{(159,\ 1.681)}{sy{=}1.403} \wr \underset{(175,\ 1.864)}{sn{=}1.777} \bullet \underset{(101,\ 2.136)}{on{=}1.861} \bullet \underset{(100,\ 2.623)}{oy{=}2.510}$
$\underline{syoy = 4.143}$ $snoy = 2.694$ 1st-2nd<4th
* FOUR left pic. # of fix.: $\underset{(168,\ 1.751)}{rs{=}1.524} \wr \underset{(166,\ 1.823)}{ls{=}1.675} \bullet \underset{(99,\ 2.107)}{lo{=}1.778} \bullet \underset{(102,\ 2.615)}{ro{=}2.578}$
$\underline{rsro = 3.966}$ $lsro = 3.328$ 1st-2nd<4th

.. TWO, $t = 2.515$ ($p < 0.05$) left pic. 1st fix. time:
$SVO = 172.614\ (334,\ 162.495) \leftarrow OVS = 211.368\ (201,\ 188.201)$
. FOUR left pic. 1st fix. time: $\underset{(159,\ 161.882)}{sy{=}160.572} \wr \underset{(175,\ 162.740)}{sn{=}183.554} \bullet \underset{(101,\ 187.074)}{on{=}202.792} \wr \underset{(100,\ 189.879)}{oy{=}220.030}$
$syoy = 2.690$ 1st<4th

** TWO, $t = 3.762$ ($p < 0.001$) left pic. 1st pass time:
$SVO = 252.790\ (334,\ 266.803) \leftarrow OVS = 371.637\ (201,\ 463.950)$
.. TWO, $t = 2.297$ ($p < 0.05$) left pic. 1st pass time:
$toLeft = 332.552\ (270,\ 440.383) \rightarrow toRight = 261.668\ (265,\ 243.751)$
* FOUR left pic. 1st pass time: $\underset{(159,\ 225.263)}{sy{=}217.943} \wr \underset{(175,\ 296.708)}{sn{=}284.451} \bullet \underset{(101,\ 281.895)}{on{=}295.861} \bullet \underset{(100,\ 585.496)}{oy{=}448.170}$
$\underline{syoy = 4.465}$ $snoy = 3.075$ 1st-2nd<4th
* FOUR left pic. 1st pass time: $\underset{(166,\ 217.223)}{ls{=}241.922} \wr \underset{(168,\ 308.361)}{rs{=}263.530} \bullet \underset{(99,\ 280.789)}{lo{=}294.778} \bullet \underset{(102,\ 581.658)}{ro{=}446.235}$
$\underline{lsro = 4.089}$ $rsro = 3.368$ 1st-2nd<4th

*. TWO, $t = 3.223$ ($p < 0.01$) left pic. 1st pass #:
$SVO = 0.994\ (334,\ 0.873) \leftarrow OVS = 1.303\ (201,\ 1.346)$
.. TWO, $t = 2.501$ ($p < 0.05$) left pic. 1st pass #:
$toLeft = 1.226\ (270,\ 1.284) \rightarrow toRight = 0.992\ (265,\ 0.821)$
* FOUR left pic. 1st pass #: $\underset{(159,\ 0.794)}{sy{=}0.874} \wr \underset{(101,\ 0.919)}{on{=}1.069} \bullet \underset{(175,\ 0.929)}{sn{=}1.103} \lhd \underset{(100,\ 1.642)}{oy{=}1.540}$
$\underline{syoy = 4.369}$ $snoy = 2.822$ 1st<4th
* FOUR left pic. 1st pass #: $\underset{(166,\ 0.778)}{ls{=}0.976} \wr \underset{(168,\ 0.960)}{rs{=}1.012} \circ \underset{(99,\ 0.892)}{lo{=}1.020} \prec \underset{(102,\ 1.632)}{ro{=}1.578}$
$\underline{lsro = 4.067}$ $\underline{rsro = 3.593}$ $loro = 2.997$

.. TWO, $t = 2.428$ ($p < 0.05$) middle pic. 1st fix. time:
$toLeft = 254.781\ (270,\ 138.453) \rightarrow toRight = 226.475\ (265,\ 131.014)$
. FOUR middle pic. 1st fix. time: $\underset{(125,\ 133.813)}{ly{=}219.888} \wr \underset{(140,\ 128.660)}{ln{=}232.357} \bullet \underset{(134,\ 137.377)}{ry{=}245.351} \wr \underset{(136,\ 139.386)}{rn{=}264.074}$
$lyrn = 2.608$ 1st<4th
. FOUR middle pic. 1st fix. time: $\underset{(99,\ 130.443)}{lo{=}214.495} \wr \underset{(166,\ 131.225)}{ls{=}233.620} \bullet \underset{(102,\ 150.558)}{ro{=}244.176} \wr \underset{(168,\ 130.606)}{rs{=}261.220}$
$lors = 2.825$ 1st<4th

** TWO, $t = 4.258$ ($p < 0.001$) middle pic. 1st pass time:
$SVO = 506.629\ (334,\ 374.245) \rightarrow OVS = 370.567\ (201,\ 329.093)$
* FOUR middle pic. 1st pass time: $\underset{(100,\ 299.832)}{oy{=}328.960} \bullet \underset{(101,\ 352.363)}{on{=}411.762} \bullet \underset{(159,\ 357.541)}{sy{=}486.447} \wr \underset{(175,\ 388.914)}{sn{=}524.966}$
$\underline{oysy = 3.667}$ $\underline{oysn = 4.353}$ 1st<3rd-4th
. FOUR middle pic. 1st pass time: $\underset{(99,\ 348.138)}{lo{=}364.283} \circ \underset{(102,\ 311.102)}{ro{=}376.667} \prec \underset{(168,\ 351.445)}{rs{=}491.030} \circ \underset{(166,\ 396.429)}{ls{=}522.416}$
$lors = 2.856$ $lols = 3.284$ $rors = 2.705$ $rols = 3.162$

** TWO, $t = 4.833$ ($p < 0.001$) middle pic. 1st pass #:
$SVO = 1.880\ (334,\ 1.174) \rightarrow OVS = 1.418\ (201,\ 0.874)$
.. TWO, $t = 1.982$ ($p < 0.05$) middle pic. 1st pass #:
$match = 1.610\ (259,\ 0.995) \leftarrow mismatch = 1.797\ (276,\ 1.173)$
.. TWO, $t = 2.123$ ($p < 0.05$) middle pic. 1st pass #:
$toLeft = 1.607\ (270,\ 0.884) \leftarrow toRight = 1.808\ (265,\ 1.266)$
* FOUR middle pic. 1st pass #: $\underset{(100,\ 0.815)}{oy{=}1.270} \bullet \underset{(101,\ 0.910)}{on{=}1.564} \bullet \underset{(159,\ 1.041)}{sy{=}1.824} \wr \underset{(175,\ 1.285)}{sn{=}1.931}$
$\underline{oysy = 4.521}$ $\underline{oysn = 4.641}$ 1st<3rd-4th
. FOUR middle pic. 1st pass #: $\underset{(134,\ 0.899)}{ry{=}1.530} \wr \underset{(136,\ 0.867)}{rn{=}1.684} \bullet \underset{(125,\ 1.087)}{ly{=}1.696} \wr \underset{(140,\ 1.403)}{ln{=}1.907}$
$ryln = 2.637$ 1st<4th
* FOUR middle pic. 1st pass #: $\underset{(102,\ 0.784)}{ro{=}1.382} \bullet \underset{(99,\ 0.961)}{lo{=}1.455} \bullet \underset{(168,\ 0.916)}{rs{=}1.744} \bullet \underset{(166,\ 1.377)}{ls{=}2.018}$
$rors = 3.318$ $\underline{rols = 4.256}$ $\underline{lols = 3.583}$ ...

*. TWO, $t = 2.592$ ($p < 0.01$) right pic. time:
$SVO = 421.174\ (334,\ 501.624) \leftarrow OVS = 562.781\ (201,\ 761.027)$
. TWO, $t = 2.745$ ($p < 0.01$) right pic. time:
$toLeft = 402.467\ (270,\ 413.507) \leftarrow toRight = 547.642\ (265,\ 762.090)$
* FOUR right pic. time: $\underset{(159,\ 396.612)}{sy{=}386.912} \wr \underset{(101,\ 442.204)}{on{=}434.059} \bullet \underset{(175,\ 580.239)}{sn{=}452.303} \wr \underset{(100,\ 968.777)}{oy{=}692.790}$
$\underline{syoy = 3.540}$ 1st<4th
* FOUR right pic. time: $\underset{(102,\ 442.048)}{ro{=}374.676} \wr \underset{(168,\ 395.582)}{rs{=}419.339} \circ \underset{(166,\ 591.134)}{ls{=}423.030} \prec \underset{(99,\ 951.738)}{lo{=}756.586}$

$\underline{rolo = 3.666}$ $\underline{rslo = 4.042}$ $\underline{lslo = 3.520}$

.. TWO, $t = 2.451$ ($p < 0.05$) right pic. # of fix.:
$SVO = 1.692\ (334,\ 2.007) \leftarrow OVS = 2.189\ (201,\ 2.658)$
. TWO, $t = 2.712$ ($p < 0.01$) right pic. # of fix.:
$toLeft = 1.615\ (270,\ 1.729) \leftarrow toRight = 2.147\ (265,\ 2.713)$
. FOUR right pic. # of fix.: $\substack{sy=1.591\\(159,\ 1.700)} \wr \substack{on=1.723\\(101,\ 1.761)} \bullet \substack{sn=1.783\\(175,\ 2.251)} \triangleleft \substack{oy=2.660\\(100,\ 3.270)}$
$syoy = 3.449$ $snoy = 2.625$ 1st<4th
. FOUR right pic. # of fix.: $\substack{rn=1.507\\(136,\ 1.628)} \wr \substack{ry=1.724\\(134,\ 1.825)} \bullet \substack{ln=2.007\\(140,\ 2.424)} \wr \substack{ly=2.304\\(125,\ 3.006)}$
$rnly = 2.691$ 1st<4th
* FOUR right pic. # of fix.: $\substack{ro=1.520\\(102,\ 1.784)} \wr \substack{rs=1.673\\(168,\ 1.697)} \circ \substack{ls=1.711\\(166,\ 2.283)} \prec \substack{lo=2.879\\(99,\ 3.192)}$
$\underline{rolo = 3.741}$ $\underline{rslo = 4.029}$ $lslo = 3.460$

.. TWO, $t = 2.108$ ($p < 0.05$) right pic. 1st pass time:
$SVO = 264.228\ (334,\ 250.216) \leftarrow OVS = 324.806\ (201,\ 414.654)$
.. TWO, $t = 2.525$ ($p < 0.05$) right pic. 1st pass time:
$toLeft = 252.237\ (270,\ 257.465) \leftarrow toRight = 322.392\ (265,\ 375.412)$
* FOUR right pic. 1st pass time: $\substack{ro=229.941\\(102,\ 298.184)} \wr \substack{ls=262.663\\(166,\ 270.553)} \circ \substack{rs=265.774\\(168,\ 229.159)} \prec \substack{lo=422.545\\(99,\ 490.143)}$
$rolo = 3.377$ $lslo = 3.421$ $\underline{rslo = 3.543}$

.. TWO, $t = 2.295$ ($p < 0.05$) right pic. 1st pass #:
$SVO = 1.069\ (334,\ 0.886) \leftarrow OVS = 1.294\ (201,\ 1.378)$
. TWO, $t = 3.037$ ($p < 0.01$) right pic. 1st pass #:
$toLeft = 1.011\ (270,\ 0.869) \leftarrow toRight = 1.298\ (265,\ 1.281)$
. FOUR right pic. 1st pass #: $\substack{sy=1.031\\(159,\ 0.860)} \wr \substack{sn=1.103\\(175,\ 0.910)} \bullet \substack{on=1.119\\(101,\ 0.993)} \wr \substack{oy=1.470\\(100,\ 1.666)}$
$syoy = 2.783$ 1st<4th
. FOUR right pic. 1st pass #: $\substack{rn=0.971\\(136,\ 0.834)} \wr \substack{ry=1.052\\(134,\ 0.904)} \bullet \substack{ln=1.243\\(140,\ 1.017)} \wr \substack{ly=1.360\\(125,\ 1.526)}$
$rnly = 2.585$ 1st<4th
* FOUR right pic. 1st pass #: $\substack{ro=0.941\\(102,\ 0.963)} \wr \substack{rs=1.054\\(168,\ 0.806)} \circ \substack{ls=1.084\\(166,\ 0.962)} \prec \substack{lo=1.657\\(99,\ 1.630)}$
$\underline{rolo = 3.802}$ $\underline{rslo = 4.034}$ $\underline{lslo = 3.595}$

.. TWO, $t = 2.564$ ($p < 0.05$) S-CHAR time:
$match = 1072.699\ (259,\ 833.868) \rightarrow mismatch = 893.014\ (276,\ 786.943)$
* FOUR S-CHAR time: $\substack{on=693.980\\(101,\ 632.761)} \prec \substack{sy=977.648\\(159,\ 642.420)} \circ \substack{sn=1007.886\\(175,\ 844.022)} \wr \substack{oy=1223.830\\(100,\ 1056.337)}$
$\underline{onsy = 3.491}$ $onsn = 3.247$ $\overline{\underline{onoy = 4.319}}$
. FOUR S-CHAR time: $\substack{rn=841.199\\(136,\ 576.478)} \wr \substack{ln=943.350\\(140,\ 947.233)} \bullet \substack{ry=1026.425\\(134,\ 793.658)} \wr \substack{ly=1122.304\\(125,\ 875.410)}$
$rnly = 3.087$ 1st<4th

*. TWO, $t = 2.698$ ($p < 0.01$) S-CHAR # of fix.:
$match = 3.869\ (259,\ 2.780) \rightarrow mismatch = 3.232\ (276,\ 2.679)$
* FOUR S-CHAR # of fix.: $\substack{on=2.436\\(101,\ 2.273)} \prec \substack{sy=3.585\\(159,\ 2.437)} \circ \substack{sn=3.691\\(175,\ 2.791)} \wr \substack{oy=4.320\\(100,\ 3.213)}$
$\underline{onsy = 3.803}$ $\underline{onsn = 3.845}$ $\overline{\underline{onoy = 4.804}}$
. FOUR S-CHAR # of fix.: $\substack{rn=3.103\\(136,\ 2.199)} \wr \substack{ln=3.357\\(140,\ 3.077)} \bullet \substack{ry=3.522\\(134,\ 2.352)} \wr \substack{ly=4.240\\(125,\ 3.143)}$

$rnly = 3.409$ 1st<4th

*. FOUR S-CHAR 1st fix. time: $\underset{(125,\ 115.896)}{ly{=}218.320} \bullet \underset{(140,\ 163.671)}{ln{=}255.986} \bullet \underset{(136,\ 146.343)}{rn{=}260.757} \wr \underset{(134,\ 154.420)}{ry{=}262.478}$
$lyrn = 2.582$ $lyry = 2.588$ 1st<3rd-4th

.. TWO, $t = 2.294$ ($p < 0.05$) S-CHAR 1st pass time:
$match = 556.270\ (259,\ 498.579) \rightarrow mismatch = 469.083\ (276,\ 375.187)$
* FOUR S-CHAR 1st pass time: $\underset{(101,\ 330.135)}{on{=}372.257} \prec \underset{(159,\ 357.541)}{sy{=}486.447} \circ \underset{(175,\ 388.914)}{sn{=}524.966} \bullet \underset{(100,\ 650.781)}{oy{=}667.290}$
$onsy = 2.585$ $onsn = 3.316$ $\underline{onoy = 4.059}$ $syoy = 2.882$ 2nd<4th

*. TWO, $t = 2.871$ ($p < 0.01$) S-CHAR 1st pass #:
$match = 2.031\ (259,\ 1.428) \rightarrow mismatch = 1.703\ (276,\ 1.212)$
* FOUR S-CHAR 1st pass #: $\underset{(101,\ 0.956)}{on{=}1.307} \prec \underset{(159,\ 1.041)}{sy{=}1.824} \circ \underset{(175,\ 1.285)}{sn{=}1.931} \bullet \underset{(100,\ 1.845)}{oy{=}2.360}$
$\underline{onsy = 4.028}$ $\underline{onsn = 4.251}$ $\underline{onoy = 5.087}$ $syoy = 2.987$ 2nd<4th
. FOUR S-CHAR 1st pass #: $\underset{(136,\ 0.976)}{rn{=}1.603} \wr \underset{(140,\ 1.400)}{ln{=}1.800} \bullet \underset{(134,\ 1.386)}{ry{=}1.940} \wr \underset{(125,\ 1.470)}{ly{=}2.128}$
$rnly = 3.425$ 1st<4th

** TWO, $t = 5.700$ ($p < 0.001$) O-CHAR time:
$SVO = 600.458\ (334,\ 564.455) \leftarrow OVS = 947.786\ (201,\ 843.266)$
* FOUR O-CHAR time: $\underset{(159,\ 489.935)}{sy{=}575.786} \circ \underset{(175,\ 625.081)}{sn{=}622.874} \prec \underset{(100,\ 831.088)}{oy{=}901.230} \circ \underset{(101,\ 856.785)}{on{=}993.881}$
$\underline{syoy = 3.965}$ $\underline{syon = 5.002}$ $snoy = 3.142$ $\underline{snon = 4.133}$
* FOUR O-CHAR time: $\underset{(168,\ 472.560)}{rs{=}597.018} \circ \underset{(166,\ 645.730)}{ls{=}603.940} \prec \underset{(102,\ 783.950)}{ro{=}889.539} \circ \underset{(99,\ 900.365)}{lo{=}1007.798}$
$\underline{rsro = 3.827}$ $\underline{rslo = 4.885}$ $lsro = 3.236$ $\underline{lslo = 4.236}$

** TWO, $t = 5.215$ ($p < 0.001$) O-CHAR # of fix.:
$SVO = 2.311\ (334,\ 2.174) \leftarrow OVS = 3.577\ (201,\ 3.440)$
* FOUR O-CHAR # of fix.: $\underset{(159,\ 1.909)}{sy{=}2.226} \circ \underset{(175,\ 2.392)}{sn{=}2.389} \prec \underset{(100,\ 3.234)}{oy{=}3.370} \circ \underset{(101,\ 3.638)}{on{=}3.782}$
$\underline{syoy = 3.579}$ $\underline{syon = 4.507}$ $snoy = 2.870$ $\underline{snon = 3.834}$
* FOUR O-CHAR # of fix.: $\underset{(168,\ 1.920)}{rs{=}2.304} \circ \underset{(166,\ 2.410)}{ls{=}2.319} \prec \underset{(102,\ 3.353)}{ro{=}3.353} \circ \underset{(99,\ 3.530)}{lo{=}3.808}$
$rsro = 3.271$ $\underline{rslo = 4.510}$ $lsro = 2.929$ $\underline{lslo = 4.073}$

.. TWO, $t = 2.298$ ($p < 0.05$) O-CHAR 1st pass time:
$match = 326.124\ (259,\ 255.638) \leftarrow mismatch = 384.891\ (276,\ 328.634)$

*. TWO, $t = 2.676$ ($p < 0.01$) O-CHAR 1st pass #:
$match = 1.274\ (259,\ 0.796) \leftarrow mismatch = 1.478\ (276,\ 0.955)$
. FOUR O-CHAR 1st pass #: $\underset{(100,\ 0.815)}{oy{=}1.270} \wr \underset{(159,\ 0.787)}{sy{=}1.277} \wr \underset{(175,\ 0.979)}{sn{=}1.429} \bullet \underset{(101,\ 0.910)}{on{=}1.564}$
$syon = 2.701$ 2nd<4th

*. TWO, $t = 3.366$ ($p < 0.01$) N-CHAR time:
$match = 191.290\ (259,\ 270.960) \leftarrow mismatch = 278.964\ (276,\ 326.892)$
. FOUR N-CHAR time: $\substack{sy=182.566\\(159,\ 266.335)} \wr \substack{oy=205.160\\(100,\ 278.946)} \bullet \substack{on=264.416\\(101,\ 297.337)} \wr \substack{sn=287.360\\(175,\ 343.339)}$
$sysn = 3.094$ 1st<4th
. FOUR N-CHAR time: $\substack{ry=166.037\\(134,\ 267.064)} \bullet \substack{ly=218.360\\(125,\ 273.555)} \bullet \substack{rn=270.529\\(136,\ 295.231)} \wr \substack{ln=287.157\\(140,\ 355.838)}$
$ryrn = 3.048$ $ryln = 3.176$ 1st<3rd-4th

*. TWO, $t = 3.421$ ($p < 0.01$) N-CHAR # of fix.:
$match = 0.799\ (259,\ 1.037) \leftarrow mismatch = 1.163\ (276,\ 1.385)$
. FOUR N-CHAR # of fix.: $\substack{sy=0.767\\(159,\ 1.008)} \bullet \substack{oy=0.850\\(100,\ 1.086)} \bullet \substack{on=1.149\\(101,\ 1.291)} \wr \substack{sn=1.171\\(175,\ 1.440)}$
$syon = 2.661$ $sysn = 2.944$ 1st<3rd-4th
. FOUR N-CHAR # of fix.: $\substack{ry=0.746\\(134,\ 1.102)} \wr \substack{ly=0.856\\(125,\ 0.965)} \bullet \substack{rn=1.081\\(136,\ 1.211)} \wr \substack{ln=1.243\\(140,\ 1.536)}$
$ryln = 3.064$ 1st<4th

*. TWO, $t = 2.850$ ($p < 0.01$) N-CHAR 1st fix. time:
$match = 112.803\ (259,\ 141.887) \leftarrow mismatch = 148.500\ (276,\ 147.451)$
. FOUR N-CHAR 1st fix. time: $\substack{sy=108.673\\(159,\ 142.224)} \wr \substack{oy=119.370\\(100,\ 141.816)} \bullet \substack{on=131.406\\(101,\ 132.737)} \wr \substack{sn=158.366\\(175,\ 154.817)}$
$sysn = 3.045$ 1st<4th
* FOUR N-CHAR 1st fix. time: $\substack{ry=92.104\\(134,\ 129.176)} \bullet \substack{ly=134.992\\(125,\ 151.758)} \bullet \substack{ln=139.600\\(140,\ 145.627)} \wr \substack{rn=157.662\\(136,\ 149.286)}$
$ryln = 2.851$ $\underline{ryrn = 3.856}$ 1st<3rd-4th

*. TWO, $t = 2.673$ ($p < 0.01$) N-CHAR 1st pass time:
$match = 147.768\ (259,\ 198.662) \leftarrow mismatch = 195.138\ (276,\ 210.524)$
. FOUR N-CHAR 1st pass time: $\substack{ry=123.470\\(134,\ 188.633)} \bullet \substack{ly=173.816\\(125,\ 206.467)} \bullet \substack{ln=192.371\\(140,\ 216.338)} \wr \substack{rn=197.985\\(136,\ 205.127)}$
$ryln = 2.805$ $ryrn = 3.106$ 1st<3rd-4th

*. TWO, $t = 2.660$ ($p < 0.01$) N-CHAR 1st pass #:
$match = 0.637\ (259,\ 0.752) \leftarrow mismatch = 0.815\ (276,\ 0.794)$
. FOUR N-CHAR 1st pass #: $\substack{ry=0.575\\(134,\ 0.760)} \wr \substack{ly=0.704\\(125,\ 0.741)} \bullet \substack{rn=0.794\\(136,\ 0.731)} \wr \substack{ln=0.836\\(140,\ 0.853)}$
$ryln = 2.671$ 1st<4th

** TWO, $t = 3.763$ ($p < 0.001$) first NP time:
$SVO = 554.305\ (334,\ 350.717) \leftarrow OVS = 705.174\ (201,\ 576.819)$
* FOUR first NP time: $\substack{sy=504.931\\(159,\ 265.171)} \bullet \substack{sn=599.166\\(175,\ 409.023)} \bullet \substack{on=696.228\\(101,\ 491.587)} \wr \substack{oy=714.210\\(100,\ 654.140)}$
$\underline{syon = 4.066}$ $\underline{syoy = 3.595}$ 1st<3rd-4th
* FOUR first NP time: $\substack{ls=543.169\\(166,\ 337.189)} \wr \substack{rs=565.310\\(168,\ 364.264)} \bullet \substack{lo=662.535\\(99,\ 609.898)} \bullet \substack{ro=746.559\\(102,\ 542.599)}$
$\underline{lsro = 3.786}$ $rsro = 3.281$ 1st-2nd<4th

** TWO, $t = 3.898$ ($p < 0.001$) first NP # of fix.:
$SVO = 2.740\ (334,\ 1.641) \leftarrow OVS = 3.408\ (201,\ 2.314)$
* FOUR first NP # of fix.: $\substack{sy=2.541\\(159,\ 1.368)} \bullet \substack{sn=2.920\\(175,\ 1.840)} \bullet \substack{on=3.337\\(101,\ 2.178)} \wr \substack{oy=3.480\\(100,\ 2.451)}$

$\underline{syon = 3.620}$ $\underline{syoy = 3.953}$ 1st<3rd-4th
. FOUR first NP # of fix.: $\underset{(168,\ 1.610)}{rs{=}2.732} \wr \underset{(166,\ 1.676)}{ls{=}2.747} \bullet \underset{(99,\ 2.237)}{lo{=}3.293} \bullet \underset{(102,\ 2.391)}{ro{=}3.520}$
$rsro = 3.231$ $lsro = 3.105$ 1st-2nd<4th

.. TWO, $t = 2.120$ ($p < 0.05$) first NP 1st fix. time:
$toLeft = 192.974\ (270,\ 86.179) \rightarrow toRight = 177.215\ (265,\ 85.752)$
. FOUR first NP 1st fix. time: $\underset{(140,\ 87.519)}{ln{=}168.879} \wr \underset{(134,\ 84.899)}{ry{=}184.507} \bullet \underset{(125,\ 83.085)}{ly{=}186.552} \wr \underset{(136,\ 86.926)}{rn{=}201.316}$
$lnrn = 3.089$ 1st<4th

*. TWO, $t = 3.267$ ($p < 0.01$) verb time:
$SVO = 428.626\ (334,\ 308.128) \leftarrow OVS = 528.766\ (201,\ 395.029)$
* FOUR verb time: $\underset{(159,\ 244.126)}{sy{=}400.132} \bullet \underset{(175,\ 355.201)}{sn{=}454.514} \bullet \underset{(101,\ 362.344)}{on{=}507.426} \wr \underset{(100,\ 426.257)}{oy{=}550.320}$
$syon = 2.852$ $\underline{syoy = 3.604}$ 1st<3rd-4th
. FOUR verb time: $\underset{(166,\ 291.180)}{ls{=}416.000} \wr \underset{(168,\ 324.395)}{rs{=}441.101} \bullet \underset{(99,\ 366.865)}{lo{=}491.444} \bullet \underset{(102,\ 419.193)}{ro{=}564.990}$
$lsro = 3.428$ $rsro = 2.719$ 1st-2nd<4th

*. TWO, $t = 2.803$ ($p < 0.01$) verb # of fix.:
$SVO = 1.946\ (334,\ 1.350) \leftarrow OVS = 2.308\ (201,\ 1.598)$
* FOUR verb # of fix.: $\underset{(159,\ 1.033)}{sy{=}1.786} \bullet \underset{(175,\ 1.573)}{sn{=}2.091} \bullet \underset{(101,\ 1.492)}{on{=}2.208} \wr \underset{(100,\ 1.700)}{oy{=}2.410}$
$syon = 2.691$ $\underline{syoy = 3.674}$ 1st<3rd-4th

** TWO, $t = 4.774$ ($p < 0.001$) adverb time:
$SVO = 419.871\ (334,\ 291.904) \leftarrow OVS = 583.458\ (201,\ 500.850)$
* FOUR adverb time: $\underset{(159,\ 278.963)}{sy{=}378.063} \bullet \underset{(175,\ 298.927)}{sn{=}457.857} \bullet \underset{(100,\ 476.816)}{oy{=}557.780} \bullet \underset{(101,\ 524.699)}{on{=}608.881}$
$\underline{syoy = 3.826}$ $\underline{syon = 4.617}$ $snon = 3.048$ ...
* FOUR adverb time: $\underset{(168,\ 296.333)}{rs{=}407.643} \bullet \underset{(166,\ 287.717)}{ls{=}432.247} \bullet \underset{(99,\ 372.458)}{lo{=}523.838} \bullet \underset{(102,\ 596.040)}{ro{=}641.324}$
$rslo = 2.808$ $\underline{rsro = 4.287}$ $\underline{lsro = 3.851}$ ...

** TWO, $t = 4.965$ ($p < 0.001$) adverb # of fix.:
$SVO = 1.946\ (334,\ 1.394) \leftarrow OVS = 2.741\ (201,\ 2.311)$
* FOUR adverb # of fix.: $\underset{(159,\ 1.300)}{sy{=}1.761} \bullet \underset{(175,\ 1.458)}{sn{=}2.114} \bullet \underset{(100,\ 2.188)}{oy{=}2.670} \bullet \underset{(101,\ 2.436)}{on{=}2.812}$
$\underline{syoy = 4.194}$ $\underline{syon = 4.523}$ $snon = 2.978$ ...
* FOUR adverb # of fix.: $\underset{(168,\ 1.359)}{rs{=}1.857} \bullet \underset{(166,\ 1.427)}{ls{=}2.036} \bullet \underset{(99,\ 1.716)}{lo{=}2.485} \bullet \underset{(102,\ 2.756)}{ro{=}2.990}$
$rslo = 3.300$ $\underline{rsro = 4.506}$ $\underline{lsro = 3.724}$ ...

** TWO, $t = 6.502$ ($p < 0.001$) second NP time:
$SVO = 624.042\ (334,\ 369.027) \leftarrow OVS = 893.408\ (201,\ 589.190)$
* FOUR second NP time: $\underset{(159,\ 310.429)}{sy{=}574.327} \circ \underset{(175,\ 410.862)}{sn{=}669.211} \prec \underset{(100,\ 532.568)}{oy{=}873.000} \circ \underset{(101,\ 642.392)}{on{=}913.614}$
$\underline{syoy = 5.701}$ $\underline{syon = 5.698}$ $\underline{snoy = 3.544}$ $\underline{snon = 3.852}$
* FOUR second NP time: $\underset{(168,\ 371.859)}{rs{=}618.881} \circ \underset{(166,\ 367.190)}{ls{=}629.265} \prec \underset{(99,\ 581.810)}{lo{=}862.515} \circ \underset{(102,\ 597.597)}{ro{=}923.392}$
$\underline{rslo = 4.173}$ $\underline{rsro = 5.163}$ $\underline{lslo = 4.001}$ $\underline{lsro = 4.993}$

** TWO, $t = 5.785$ ($p < 0.001$) second NP # of fix.:
$SVO = 2.796\ (334,\ 1.796) \leftarrow OVS = 3.876\ (201,\ 2.504)$
* FOUR second NP # of fix.: $\underset{(159,\ 1.386)}{sy{=}2.547} \circ \underset{(175,\ 2.079)}{sn{=}3.023} \prec \underset{(101,\ 2.481)}{on{=}3.842} \circ \underset{(100,\ 2.539)}{oy{=}3.910}$
$\underline{syon = 5.391}$ $\underline{syoy = 5.579}$ $snon = 2.933$ $snoy = 3.136$
* FOUR second NP # of fix.: $\underset{(168,\ 1.793)}{rs{=}2.780} \circ \underset{(166,\ 1.804)}{ls{=}2.813} \prec \underset{(99,\ 2.217)}{lo{=}3.677} \circ \underset{(102,\ 2.751)}{ro{=}4.069}$
$\underline{rslo = 3.611}$ $\underline{rsro = 4.660}$ $lslo = 3.455$ $\underline{lsro = 4.511}$

** TWO, $t = 7.303$ ($p < 0.001$) second NP r-path time:
$SVO = 968.578\ (334,\ 799.586) \leftarrow OVS = 1678.975\ (201,\ 1448.955)$
* FOUR second NP r-path time: $\underset{(159,\ 604.260)}{sy{=}819.396} \prec \underset{(175,\ 923.684)}{sn{=}1104.120} \prec \underset{(100,\ 1451.653)}{oy{=}1651.450} \circ \underset{(101,\ 1453.001)}{on{=}1706.228}$
$sysn = 3.298$ $\underline{syoy = 6.404}$ $\underline{syon = 6.828}$ $\underline{snoy = 3.818}$ $\underline{snon = 4.206}$
* FOUR second NP r-path time: $\underset{(168,\ 837.631)}{rs{=}958.911} \circ \underset{(166,\ 761.563)}{ls{=}978.361} \prec \underset{(99,\ 1379.739)}{lo{=}1556.051} \circ \underset{(102,\ 1510.314)}{ro{=}1798.284}$
$\underline{rslo = 4.402}$ $\underline{rsro = 5.872}$ $\underline{lslo = 4.391}$ $\underline{lsro = 5.886}$

** TWO, $t = 6.811$ ($p < 0.001$) second NP r-path #:
$SVO = 4.380\ (334,\ 3.868) \leftarrow OVS = 7.373\ (201,\ 6.298)$
* FOUR second NP r-path #: $\underset{(159,\ 2.754)}{sy{=}3.654} \prec \underset{(175,\ 4.564)}{sn{=}5.040} \prec \underset{(100,\ 6.257)}{oy{=}7.370} \circ \underset{(101,\ 6.370)}{on{=}7.376}$
$sysn = 3.319$ $\underline{syoy = 6.553}$ $\underline{syon = 6.481}$ $\underline{snoy = 3.546}$ $\underline{snon = 3.531}$
* FOUR second NP r-path #: $\underset{(168,\ 3.959)}{rs{=}4.310} \circ \underset{(166,\ 3.784)}{ls{=}4.452} \prec \underset{(99,\ 5.449)}{lo{=}6.828} \circ \underset{(102,\ 7.011)}{ro{=}7.902}$
$\underline{rslo = 4.353}$ $\underline{rsro = 5.380}$ $\underline{lslo = 4.179}$ $\underline{lsro = 5.225}$

### B.4.2 Effects in *incorrectly answered* trials

.. TWO, $t = 2.510$ ($p < 0.05$) background time:
$toLeft = 244.043\ (92,\ 296.017) \rightarrow toRight = 150.926\ (94,\ 202.173)$
. FOUR background time: $\underset{(78,\ 208.147)}{lo{=}141.385} \wr \underset{(16,\ 168.035)}{ls{=}197.438} \bullet \underset{(77,\ 306.345)}{ro{=}226.104} \wr \underset{(15,\ 221.935)}{rs{=}336.133}$
$lors = 3.284$ 1st<4th

*. TWO, $t = 3.074$ ($p < 0.01$) background # of fix.:
$toLeft = 1.370\ (92,\ 1.238) \rightarrow toRight = 0.851\ (94,\ 1.057)$
. FOUR background # of fix.: $\underset{(42,\ 0.857)}{ln{=}0.738} \bullet \underset{(52,\ 1.195)}{ly{=}0.942} \bullet \underset{(45,\ 1.036)}{rn{=}1.289} \wr \underset{(47,\ 1.411)}{ry{=}1.447}$
$lnrn = 2.691$ $lnry = 2.822$ 1st<3rd-4th
* FOUR background # of fix.: $\underset{(78,\ 1.129)}{lo{=}0.846} \wr \underset{(16,\ 0.619)}{ls{=}0.875} \bullet \underset{(77,\ 1.183)}{ro{=}1.247} \bullet \underset{(15,\ 1.363)}{rs{=}2.000}$
$\underline{lors = 3.504}$ $lsrs = 2.992$ 1st-2nd<4th

*. TWO, $t = 2.924$ ($p < 0.01$) background 1st fix. time:
$SVO = 147.194\ (31,\ 118.554) \rightarrow OVS = 91.826\ (155,\ 91.284)$
. FOUR background 1st fix. time: $\underset{(78,\ 88.749)}{lo{=}75.487} \wr \underset{(77,\ 91.396)}{ro{=}108.377} \bullet \underset{(15,\ 82.853)}{rs{=}132.933} \wr \underset{(16,\ 145.977)}{ls{=}160.562}$
$lols = 3.090$ 1st<4th

*. FOUR agent 1st fix. time: $\substack{ls=122.188\\(16,\ 129.245)} \wr \substack{ro=129.636\\(77,\ 115.540)} \wr \substack{lo=156.756\\(78,\ 150.084)} \bullet \substack{rs=219.600\\(15,\ 158.261)}$
$rors = 2.588$ 2nd<4th

.. TWO, $t = 2.388$ $(p < 0.05)$ RAW middle pic. 1st fix. time:
$toLeft = 201.261\ (92,\ 93.574) \rightarrow toRight = 175.628\ (94,\ 45.067)$

** TWO, $t = 3.580$ $(p < 0.001)$ patient time:
$match = 367.000\ (99,\ 419.036) \leftarrow mismatch = 718.655\ (87,\ 869.319)$
* FOUR patient time: $\substack{sn=198.000\\(10,\ 178.762)} \wr \substack{oy=331.141\\(78,\ 369.592)} \wr \substack{sy=500.190\\(21,\ 557.543)} \bullet \substack{on=786.273\\(77,\ 900.592)}$
$\underline{oyon = 4.125}$ 2nd<4th
. FOUR patient time: $\substack{ly=364.596\\(52,\ 435.883)} \wr \substack{ry=369.660\\(47,\ 404.262)} \bullet \substack{rn=606.800\\(45,\ 611.714)} \bullet \substack{ln=838.500\\(42,\ 1074.775)}$
$lyln = 2.901$ $ryln = 2.780$ 1st-2nd<4th

** TWO, $t = 3.852$ $(p < 0.001)$ patient # of fix.:
$match = 1.323\ (99,\ 1.483) \leftarrow mismatch = 2.632\ (87,\ 2.989)$
* FOUR patient # of fix.: $\substack{sn=1.100\\(10,\ 0.738)} \wr \substack{oy=1.205\\(78,\ 1.283)} \wr \substack{sy=1.762\\(21,\ 2.047)} \bullet \substack{on=2.831\\(77,\ 3.114)}$
$\underline{oyon = 4.261}$ 2nd<4th
. FOUR patient # of fix.: $\substack{ly=1.288\\(52,\ 1.513)} \bullet \substack{ry=1.362\\(47,\ 1.466)} \bullet \substack{rn=2.333\\(45,\ 2.384)} \bullet \substack{ln=2.952\\(42,\ 3.527)}$
$lyrn = 2.612$ $lyln = 3.073$ $ryln = 2.832$ ...

.. TWO, $t = 1.996$ $(p < 0.05)$ patient 1st fix. time:
$match = 195.040\ (99,\ 170.831) \leftarrow mismatch = 249.966\ (87,\ 204.316)$

.. TWO, $t = 2.312$ $(p < 0.05)$ patient 1st pass time:
$match = 252.859\ (99,\ 270.122) \leftarrow mismatch = 344.011\ (87,\ 266.114)$
. FOUR patient 1st pass time: $\substack{sn=198.000\\(10,\ 178.762)} \wr \substack{oy=236.487\\(78,\ 241.784)} \wr \substack{sy=313.667\\(21,\ 357.123)} \bullet \substack{on=362.974\\(77,\ 270.520)}$
$oyon = 3.070$ 2nd<4th

*. TWO, $t = 2.852$ $(p < 0.01)$ patient 1st pass #:
$match = 0.929\ (99,\ 0.906) \leftarrow mismatch = 1.287\ (87,\ 0.791)$
* FOUR patient 1st pass #: $\substack{oy=0.872\\(78,\ 0.762)} \wr \substack{sn=1.100\\(10,\ 0.738)} \bullet \substack{sy=1.143\\(21,\ 1.315)} \wr \substack{on=1.312\\(77,\ 0.799)}$
$\underline{oyon = 3.508}$ 1st<4th
. FOUR patient 1st pass #: $\substack{ry=0.915\\(47,\ 0.686)} \wr \substack{ly=0.942\\(52,\ 1.074)} \bullet \substack{ln=1.238\\(42,\ 0.726)} \wr \substack{rn=1.333\\(45,\ 0.853)}$
$ryrn = 2.598$ 1st<4th

*. FOUR left pic. time: $\substack{rn=335.756\\(45,\ 537.396)} \wr \substack{ly=364.596\\(52,\ 435.883)} \bullet \substack{ry=513.809\\(47,\ 736.431)} \bullet \substack{ln=838.500\\(42,\ 1074.775)}$
$rnln = 2.787$ $lyln = 2.901$ 1st-2nd<4th

*. FOUR left pic. # of fix.: $\substack{ly=1.288\\(52,\ 1.513)} \wr \substack{rn=1.400\\(45,\ 1.763)} \bullet \substack{ry=1.894\\(47,\ 2.389)} \bullet \substack{ln=2.952\\(42,\ 3.527)}$
$lyln = 3.073$ $rnln = 2.623$ 1st-2nd<4th

*. TWO, $t = 3.043$ ($p < 0.01$) left pic. 1st fix. time:
$toLeft = 144.304\ (92,\ 126.960) \leftarrow toRight = 210.989\ (94,\ 168.554)$
* FOUR left pic. 1st fix. time: $\overset{rn=120.556}{(45,\ 114.633)} \wr \overset{ry=167.043}{(47,\ 135.048)} \bullet \overset{ly=184.231}{(52,\ 170.715)} \wr \overset{ln=244.119}{(42,\ 161.736)}$
$\underline{rnln = 4.133}$ 1st<4th
* FOUR left pic. 1st fix. time: $\overset{ro=129.636}{(77,\ 115.540)} \bullet \overset{ls=207.250}{(16,\ 179.445)} \bullet \overset{lo=211.756}{(78,\ 167.444)} \wr \overset{rs=219.600}{(15,\ 158.261)}$
$\underline{rolo = 3.549}$ $rors = 2.588$ 1st<3rd-4th

*. FOUR middle pic. 1st pass time: $\overset{rs=332.867}{(15,\ 267.771)} \wr \overset{ro=453.104}{(77,\ 380.773)} \bullet \overset{lo=470.346}{(78,\ 380.184)} \bullet \overset{ls=737.375}{(16,\ 431.483)}$
$rsls = 3.111$ $rols = 2.656$ 1st-2nd<4th

*. FOUR right pic. time: $\overset{ln=291.286}{(42,\ 380.693)} \wr \overset{ry=369.660}{(47,\ 404.262)} \bullet \overset{ly=399.442}{(52,\ 482.836)} \wr \overset{rn=606.800}{(45,\ 611.714)}$
$lnrn = 2.864$ 1st<4th

*. FOUR right pic. # of fix.: $\overset{ln=1.095}{(42,\ 1.462)} \wr \overset{ry=1.362}{(47,\ 1.466)} \bullet \overset{ly=1.577}{(52,\ 1.944)} \wr \overset{rn=2.333}{(45,\ 2.384)}$
$lnrn = 2.895$ 1st<4th

*. TWO, $t = 3.031$ ($p < 0.01$) right pic. 1st fix. time:
$toLeft = 230.685\ (92,\ 207.801) \rightarrow toRight = 150.872\ (94,\ 146.679)$
. FOUR right pic. 1st fix. time: $\overset{ln=141.214}{(42,\ 158.668)} \wr \overset{ly=158.673}{(52,\ 137.313)} \bullet \overset{ry=207.000}{(47,\ 171.997)} \wr \overset{rn=255.422}{(45,\ 239.069)}$
$lnrn = 2.606$ 1st<4th
. FOUR right pic. 1st fix. time: $\overset{ls=122.188}{(16,\ 129.245)} \wr \overset{lo=156.756}{(78,\ 150.084)} \wr \overset{rs=161.600}{(15,\ 117.588)} \bullet \overset{ro=244.143}{(77,\ 219.161)}$
$loro = 2.900$ 2nd<4th

*. FOUR right pic. 1st pass #: $\overset{ln=0.905}{(42,\ 1.100)} \wr \overset{ry=0.915}{(47,\ 0.686)} \wr \overset{ly=1.173}{(52,\ 1.294)} \bullet \overset{rn=1.333}{(45,\ 0.853)}$
$ryrn = 2.598$ 2nd<4th

*. TWO, $t = 2.984$ ($p < 0.01$) S-CHAR time:
$SVO = 1117.871\ (31,\ 866.530) \rightarrow OVS = 641.342\ (155,\ 800.682)$
* FOUR S-CHAR time: $\overset{oy=498.269}{(78,\ 663.256)} \wr \overset{sn=720.700}{(10,\ 623.186)} \bullet \overset{on=786.273}{(77,\ 900.592)} \wr \overset{sy=1307.000}{(21,\ 913.848)}$
$\underline{oysy = 4.556}$ 1st<4th
* FOUR S-CHAR time: $\overset{ro=598.169}{(77,\ 719.976)} \wr \overset{lo=683.962}{(78,\ 875.696)} \bullet \overset{rs=929.000}{(15,\ 976.392)} \bullet \overset{ls=1294.938}{(16,\ 736.782)}$
$\underline{rols = 3.509}$ $lols = 2.605$ 1st-2nd<4th

*. TWO, $t = 2.786$ ($p < 0.01$) S-CHAR # of fix.:
$SVO = 4.000\ (31,\ 3.975) \rightarrow OVS = 2.355\ (155,\ 2.772)$
* FOUR S-CHAR # of fix.: $\overset{oy=1.885}{(78,\ 2.313)} \wr \overset{sn=2.800}{(10,\ 3.155)} \bullet \overset{on=2.831}{(77,\ 3.114)} \wr \overset{sy=4.571}{(21,\ 4.261)}$
$\underline{oysy = 3.866}$ 1st<4th

*. TWO, $t = 2.592$ ($p < 0.01$) S-CHAR 1st fix. time:
$match = 192.222\ (99,\ 176.533) \leftarrow mismatch = 264.690\ (87,\ 204.776)$
* FOUR S-CHAR 1st fix. time: $\overset{oy=164.564}{(78,\ 130.874)} \bullet \overset{sn=261.900}{(10,\ 164.258)} \bullet \overset{on=265.052}{(77,\ 210.368)} \wr \overset{sy=294.952}{(21,\ 269.621)}$

$\underline{oyon = 3.576}$ $oysy = 3.137$ 1st<3rd-4th
. FOUR S-CHAR 1st fix. time: $\overset{ry=175.489}{(47, 132.273)} \wr \overset{ly=207.346}{(52, 208.833)} \bullet \overset{rn=259.378}{(45, 239.796)} \wr \overset{ln=270.381}{(42, 161.814)}$
$ryln = 3.041$ 1st<4th

*. TWO, $t = 3.223$ ($p < 0.01$) S-CHAR 1st pass time:
$SVO = 541.645\ (31,\ 410.824) \rightarrow OVS = 338.271\ (155,\ 300.089)$
. FOUR S-CHAR 1st pass time: $\overset{oy=313.885}{(78, 326.594)} \wr \overset{on=362.974}{(77, 270.520)} \bullet \overset{sn=403.500}{(10, 301.691)} \bullet \overset{sy=607.429}{(21, 445.107)}$
$oysy = 3.370$ $onsy = 3.153$ 1st-2nd<4th
* FOUR S-CHAR 1st pass time: $\overset{ro=330.675}{(77, 307.845)} \wr \overset{rs=332.867}{(15, 267.771)} \circ \overset{lo=345.769}{(78, 294.031)} \prec \overset{ls=737.375}{(16, 431.483)}$
$\underline{rols = 4.466}$ $rsls = 3.111$ $\underline{lols = 4.452}$

*. FOUR S-CHAR 1st pass #: $\overset{rs=1.200}{(15, 0.561)} \wr \overset{ro=1.273}{(77, 0.968)} \wr \overset{lo=1.321}{(78, 1.111)} \bullet \overset{ls=2.125}{(16, 1.408)}$
$rols = 2.944$ 2nd<4th

*. TWO, $t = 3.024$ ($p < 0.01$) O-CHAR time:
$SVO = 512.097\ (31,\ 560.433) \leftarrow OVS = 1005.555\ (155,\ 872.143)$
.. TWO, $t = 2.335$ ($p < 0.05$) O-CHAR time:
$match = 788.919\ (99,\ 743.664) \leftarrow mismatch = 1076.241\ (87,\ 932.802)$
. FOUR O-CHAR time: $\overset{sy=500.190}{(21, 557.543)} \wr \overset{sn=537.100}{(10, 595.933)} \bullet \overset{oy=866.654}{(78, 770.887)} \wr \overset{on=1146.260}{(77, 948.335)}$
$syon = 2.978$ 1st<4th

*. TWO, $t = 3.243$ ($p < 0.01$) O-CHAR # of fix.:
$SVO = 1.871\ (31,\ 2.029) \leftarrow OVS = 3.503\ (155,\ 2.649)$
.. TWO, $t = 2.376$ ($p < 0.05$) O-CHAR # of fix.:
$match = 2.808\ (99,\ 2.436) \leftarrow mismatch = 3.713\ (87,\ 2.757)$
. FOUR O-CHAR # of fix.: $\overset{sy=1.762}{(21, 2.047)} \wr \overset{sn=2.100}{(10, 2.079)} \bullet \overset{oy=3.090}{(78, 2.466)} \wr \overset{on=3.922}{(77, 2.776)}$
$syon = 3.323$ 1st<4th

*. TWO, $t = 2.710$ ($p < 0.01$) N-CHAR 1st fix. time:
$match = 183.566\ (99,\ 168.358) \rightarrow mismatch = 123.115\ (87,\ 130.431)$
. FOUR N-CHAR 1st fix. time: $\overset{on=121.727}{(77, 135.119)} \wr \overset{sn=133.800}{(10, 90.817)} \bullet \overset{sy=155.524}{(21, 155.252)} \wr \overset{oy=191.115}{(78, 171.878)}$
$onoy = 2.792$ 1st<4th
. FOUR N-CHAR 1st fix. time: $\overset{rn=115.311}{(45, 106.565)} \wr \overset{ln=131.476}{(42, 152.837)} \bullet \overset{ly=150.192}{(52, 154.122)} \wr \overset{ry=220.489}{(47, 177.196)}$
$rnry = 3.431$ 1st<4th

*. TWO, $t = 2.617$ ($p < 0.01$) first NP time:
$match = 474.020\ (99,\ 299.732) \leftarrow mismatch = 616.115\ (87,\ 435.577)$

*. TWO, $t = 2.711$ ($p < 0.01$) first NP # of fix.:
$match = 2.374\ (99,\ 1.556) \leftarrow mismatch = 3.103\ (87,\ 2.102)$

.. TWO, $t = 2.502$ ($p < 0.05$) first NP 1st fix. time:
$SVO = 145.806\ (31,\ 92.833) \leftarrow OVS = 189.865\ (155,\ 88.850)$

*. TWO, $t = 2.970$ ($p < 0.01$) verb 1st fix. time:
$SVO = 166.935\ (31,\ 93.983) \leftarrow OVS = 218.613\ (155,\ 87.311)$
. FOUR verb 1st fix. time: $\substack{sn=147.400\\(10,\ 117.405)} \wr \substack{sy=176.238\\(21,\ 82.247)} \bullet \substack{on=209.052\\(77,\ 85.885)} \wr \substack{oy=228.051\\(78,\ 88.230)}$
$snoy = 2.618$ 1st<4th
. FOUR verb 1st fix. time: $\substack{rs=156.400\\(15,\ 70.073)} \wr \substack{ls=176.812\\(16,\ 113.435)} \bullet \substack{ro=213.182\\(77,\ 85.879)} \wr \substack{lo=223.974\\(78,\ 88.930)}$
$rslo = 2.777$ 1st<4th

.. TWO, $t = 1.982$ ($p < 0.05$) verb 1st pass time:
$SVO = 184.290\ (31,\ 145.318) \leftarrow OVS = 239.181\ (155,\ 139.873)$

*. FOUR adverb 1st fix. time: $\substack{rs=149.133\\(15,\ 96.157)} \wr \substack{ro=214.883\\(77,\ 103.569)} \bullet \substack{ls=228.688\\(16,\ 89.126)} \wr \substack{lo=229.962\\(78,\ 94.996)}$
$rslo = 3.012$ 1st<4th

# References

[AK99] Gerry T. M. Altmann and Yuki Kamide. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73:247 – 264, 1999.

[Alt98] Gerry T.M. Altmann. Ambiguity in sentence processing. *Trends in Cognitive Sciences*, 2(4):146 – 152, 1998.

[Aue05] Eric Auer. *NEWTRACK – reading experiments with the Fourward DPI eye-tracker on modern PCs*, 2004 – 2005. http://www.coli.uni-saarland.de/~eric/newtrack/.

[CJ75] Patricia A. Carpenter and Marcel Adam Just. Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 82(1):45 – 73, 1975.

[CYG92] P. J. Carroll, J. R. Young, and M. S. Guertin. Visual analysis of cartoons: A view from the far side. In K. Rayner, editor, *Eye movements and visual cognition: Scene perception and reading*, pages 444 – 461. Springer Verlag, 1992.

[Heg92] M. Hegarty. The mechanics of comprehension and comprehension of mechanics. In K. Rayner, editor, *Eye movements and visual cognition: Scene perception and reading*, pages 428 – 443. Springer Verlag, 1992.

[Hem93] Barbara Hemforth. *Kognitives Parsing: Repräsentation und Verarbeitung sprachlichen Wissens.* Infix, Sankt Augustin, 1993.

[KAH03] Yuki Kamide, Gerry T. M. Altmann, and Sarah L. Haywood. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 2003. http://www.stanford.edu/~jurafsky/kamide.pdf.

[Kas03] Thorsten Kaske. *AscFilter User Manual*, 2003.

[KC04] Pia Knoeferle and Matthew W. Crocker. The coordinated processing of scene and utterance: evidence from eye-tracking in depicted events, 2004.

[KC05] Pia Knoeferle and Matthew W. Crocker. Incremental effects of mismatch during picture-sentence integration: Evidence from eye-tracking, 2005.

[KCSP05] Pia Knoeferle., Matthew W. Crocker, Christoph Scheepers, and Martin J. Pickering. The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition*, 95:95 – 127, 2005.

[Kno05] Pia Knoeferle. *The role of visual scenes in spoken language comprehension: Evidence from eye-tracking.* Doctoral dissertation in computational linguistics, Saarland University, http://scidok.sulb.uni-saarland.de/volltexte/2005/438/, 2005.

[KSA03] Yuki Kamide, Christoph Scheepers, and Gerry T. M. Altmann. Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from german and english. *Journal of Psycholinguistic Research*, 32:37 – 55, 2003.

[Ray98] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124:372 – 422, 1998.

[RRS+01] Keith Rayner, Caren M. Rotello, Andrew J. Stewart, Jessica Keir, and Susan A. Duffy. Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology*, 7(3):219 – 226, 2001.

[SBC98] Christoph Scheepers, Holly Branigan, and Siobhan Campbell. Data analysis (ii): Eye-tracking measures (v 1.0). EyeGaze, 8 1998.

[STCC99] Julie C. Sedivy, Michael K. Tanenhaus, C. G. Chambers, and G. N. Carlson. Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71:109 – 148, 1999. http://www.bcs.rochester.edu/people/mtan/sedivy1999.pdf.

[TSKES95] Michael K. Tanenhaus, Michael J. Spivey-Knowlton, Kathleen M. Eberhard, and Julie C. Sedivy. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268:632 – 634, 1995. http://www.bcs.rochester.edu/people/mtan/mtan1995.pdf.

[Tuk77] John W. Tukey. *Exploratory Data Analysis.* Addison-Weslay, Reading (MA), 1977.

[UJR04] Geoffrey Underwood, Lorraine Jebbet, and Katharine Roberts. Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *The Quarterly Journal of Experimental Psychology*, 57(1):165 – 182, 2004.

[VZBdL99] Troy A. W. Visser, Samantha M. Zuvic, Walter F. Bischof, and Vincent di Lollo. The attentional blink with targets in different spatial locations. *Psychonomic Bulletin & Review*, 6(3):432 – 436, 1999.

[@BORLAND] Borland Developer Network Museum:
http://community.borland.com/museum/

[@DJGPP] DJGPP Compiler by DJ Delorie, a DOS version of GCC/GPP:
http://www.delorie.com/djgpp/

[@EYELINK] EyeLink II by SR Research (head-mounted camera eyetrackers):
http://www.eyelinkinfo.com/

[@FREEDOS] FreeDOS, a DOS operating system:
http://www.freedos.org/

[@GPL] GNU General Public License (by the Free Software Foundation, Inc.):
http://www.gnu.org/licenses/gpl.html

[@NASM] NASM Net-Wide Assembler:
http://nasm.sourceforge.net/

[@NYU] Study Guide for Essentials of Statistics for the Social and Behavioral Sciences, by Barry H. Cohen and R. Brooke Lea:
http://www.psych.nyu.edu/cohen/StudyGuide.pdf

[@PCEXPT] PCEXPT eyetracking software by Charles Clifton, Jr.:
http://www-unix.oit.umass.edu/~cec/software.html

[@RBIL] Ralf Brown's Files (RBdualVGA, RBIL, others):
http://www-2.cs.cmu.edu/afs/cs/user/ralf/pub/WWW/files.html

[@SDD] SciTech Display Doctor (free for DOS VESA VBE3 driver):
http://www.scitechsoft.com/products/ent/free_titles.html

[@SMI] SensoMotoric Instruments GmbH (head-mounted camera eyetrackers):
http://www.smi.de/

[@VBEHZ] VBEHz, a free tool to change your screen refresh rate:
http://home.arcor.de/g.s/vbehz.htm

[@WARD] Fourward Technologies, Inc. (DPI eyetracker):
http://www.fourward.com/

[@WOLFRAM] Wolfram Research MathWorld – Eric W. Weisstein):
http://mathworld.wolfram.com/



Statutory declaration: I hereby declare to have written this thesis on my own, only using the resources specified in this thesis. (Eidesstattliche Erklärung: Hiermit erkläre ich, daß ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Hilfsmittel angefertigt habe.)

Saarbrücken, 6th September 2005 Eric Auer