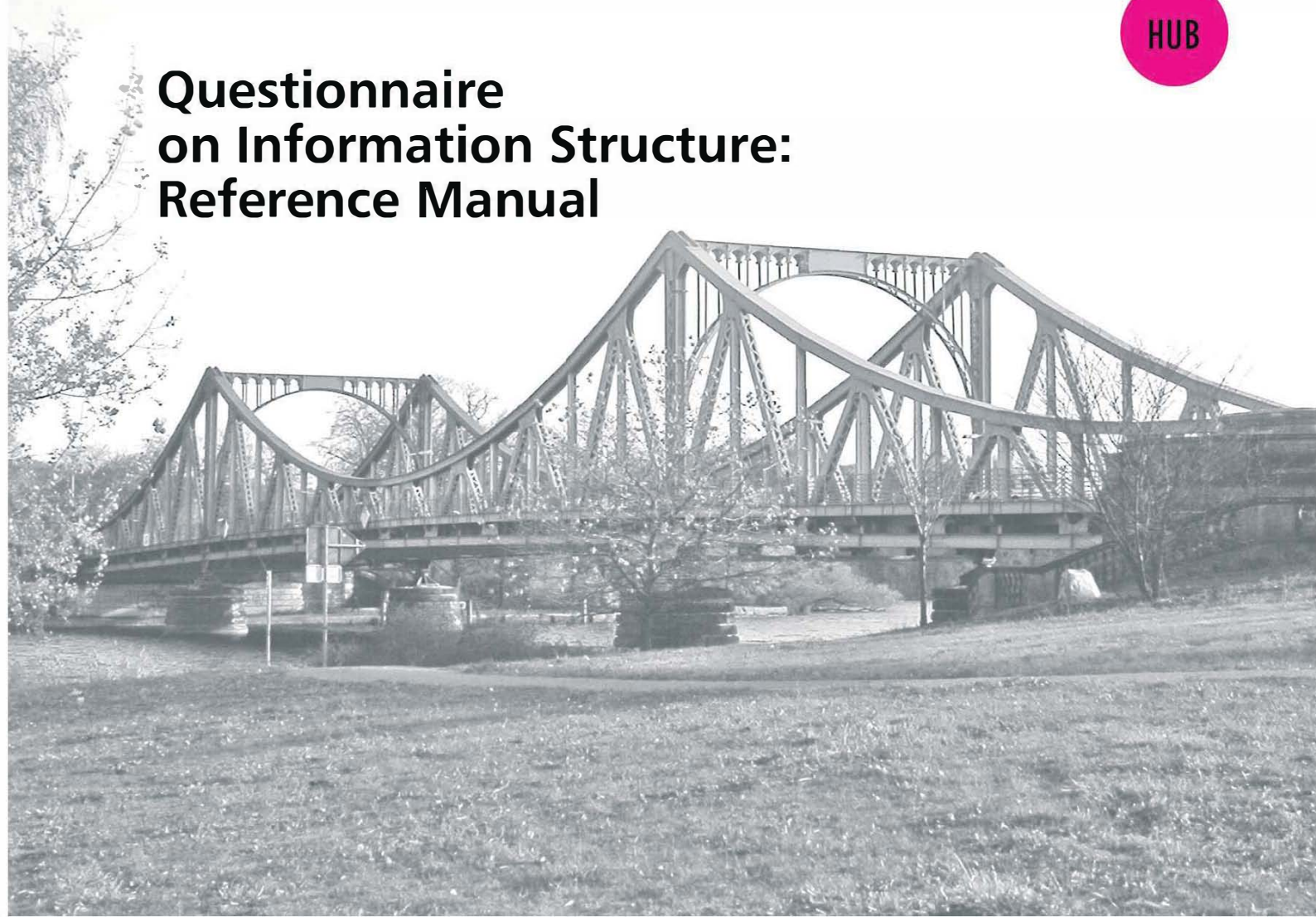


Interdisciplinary Studies on Information Structure Vol. 4

UP

Questionnaire on Information Structure: Reference Manual

HUB



Stavros Skopeteas

Ines Fiedler

Sam Hellmuth

Anne Schwarz

Ruben Stoel

Gisbert Fanselow

Caroline Féry

Manfred Krifka

Questionnaire on Information Structure (QUIS):
Reference Manual

Bibliografische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

ISIS issues do not appear according to strict schedule.

© Copyrights of articles remain with the authors.

Vol. 4 (2006)

Volume Editors: Stavros Skopeteas¹, Ines Fiedler²
Samantha Hellmuth¹, Anne Schwarz², Ruben Stoel¹
Gisbert Fanselow¹, Caroline Féry¹, Manfred Krifka²
¹Universität Potsdam, SFB 632
Postfach 60 15 53, D-14415 Potsdam
²Humboldt-Universität zu Berlin, SFB 632
Sitz: Mohrenstr. 40-41
Unter den Linden 6, D-10099 Berlin

Series Editors: Shinichiro Ishihara, Michaela Schmitz
Universität Potsdam, SFB 632
Postfach 60 15 53, D-14415 Potsdam
ishihara@uni-potsdam.de; mschmitz@ling.uni-potsdam.de

Anne Schwarz
Humboldt-Universität zu Berlin, SFB 632
Sitz: Mohrenstr. 40-41
Unter den Linden 6, D-10099 Berlin
anne.schwarz@rz.hu-berlin.de

Published by Universitätsverlag Potsdam
Postfach 60 15 53, D-14415 Potsdam
Fon +49 (0) 331 977 4517
Fax +49 (0) 331 977 4625
e-mail: ubpub@uni-potsdam.de
<http://info.ub.uni-potsdam.de/verlag.htm>

Printed by Audiovisuelles Zentrum der Universität Potsdam
und sd:k Teltow

Published 2006

ISSN **1614-4708**
ISBN **3-939469-14-9**
ISBN **978-3-939469-14-8**

Contents

Contents	iii
Members of project D2 “Typology of information structure”	v
Chapter 1. Introduction	1
1 Information Structure.....	1
2 Grammatical Correlates of Information Structure	4
3 Structure of the Questionnaire	5
4 Experimental Tasks.....	7
5 Technicalities	11
6 Archiving	11
7 Acknowledgments.....	12
Chapter 2. General Questions	13
1 General Information.....	13
2 Phonology	14
3 Morphology and Syntax.....	18
Chapter 3. Experimental tasks	31
1 Changes (Given/New in Intransitives and Transitives).....	31
2 Giving (Given/New in Ditransitives).....	36
3 Visibility (Given/New, Animacy and Type/Token Reference).....	39
4 Locations (Given/New in Locative Expressions).....	73
5 Sequences (Given/New/Contrast in Transitives).....	77
6 Dynamic Localization (Given/New in Dynamic Loc. Descriptions)	82
7 Birthday Party (Weight and Discourse Status).....	85
8 Static Localization (Macro-Planning and Given/New in Locatives).....	93
9 Guiding (Presentational Utterances).....	95
10 Event Cards (All New).....	98
11 Anima (Focus types and Animacy).....	99
12 Contrast (Contrast in pairing events).....	107
13 Animal Game (Broad/Narrow Focus in NP)	111
14 Properties (Focus on Property and Possessor).....	117

15	Eventives (Thetic and Categorical Utterances)	122
16	Tell a Story (Contrast in Text)	126
17	Focus Cards (Selective, Restrictive, Additive, Rejective Focus)	137
18	Who does What (Answers to Multiple Constituent Questions)	142
19	Fairy Tale (Topic and Focus in Coherent Discourse).....	149
20	Map Task (Contrastive and Selective Focus in Spontaneous Dialogue)	155
21	Drama (Contrastive Focus in Argumentation).....	159
22	Events in Places (Spatial, Temporal and Complex Topics)	161
23	Path Descriptions (Topic Change in Narrative).....	167
24	Groups (Partial Topic)	169
25	Connections (Bridging Topic)	174
26	Indirect (Implicational Topic).....	185
27	Surprises (Subject-Topic Interrelation)	194
28	Doing (Action Given, Action Topic)	196
29	Influences (Question Priming).....	201
Chapter 4. Translation tasks		205
1	Basic Intonational Properties	205
2	Focus Translation.....	209
3	Topic Translation	225
4	Quantifiers.....	228
Chapter 5. Information structure summary survey		231
1	Preliminaries	231
2	Syntax.....	232
3	Morphology.....	237
4	Prosody.....	238
5	Summary: Information structure.....	247
Chapter 6. Performance of Experimental Tasks in the Field		249
1	Field sessions	249
2	Field Session Metadata	251
3	Informants' Agreement.....	255
References		257

Members of project D2 “Typology of information structure”

Staff

Gisbert Fanselow	fanselow@rz.uni-potsdam.de
Caroline Féry	caroline.fery@gmail.com
Manfred Krifka	krifka@rz.hu-berlin.de
Ines Fiedler	ines.fiedler@staff.hu-berlin.de
Sam Hellmuth (since April 2006)	sam.hellmuth@rz.uni-potsdam.de
Anne Schwarz	anne.schwarz@rz.hu-berlin.de
Stavros Skopeteas	stavros.skopeteas@rz.uni-potsdam.de
Ruben Stoel (until December 2005)	r.b.stoel@let.leidenuniv.nl

Technical and scientific assistance

Anja Arnhold (until August 2005)	arnhold@ling.uni-potsdam.de
Elizabeth Medvedovsky (until April 2006)	elizabeth@ling.uni-potsdam.de
Katharina Moczko (since September 2005)	kmoczko@web.de
Andreas Pankau	pankau@ling.uni-potsdam.de
Fabian Schubö (since October 2005)	fabian.schuboe@web.de
Bei Wang	bwang@rz.uni-potsdam.de

Sponsored by the *German Research Foundation* (DFG) as part of the SFB 632 “Information Structure” (University of Potsdam/Humboldt University Berlin).

Chapter 1. Introduction

This bulky manual contains a questionnaire for the investigation of information structure from a typological perspective. It provides a tool for the collection of natural linguistic data, both spoken and written, and, secondly, for the elaboration of grammars of information structure in genetically diverse languages.

1 Information Structure

Information structure is concerned both with ‘mental states’ of speakers and hearers and with linguistic means used to convey these mental states. In other words, the linguist interested in information structure (IS), deals simultaneously with formal and communicative aspects of language. The main contrasts concern ‘new’, ‘accessible’ and ‘given’, as well as ‘topic’, focus,’ and ‘background’, though finer divisions are also used below. A focus for instance can be wide or narrow, it can be ‘out of the blue’, informational, contrastive, selective or corrective, etc. For Clark & Haviland (1977:3), *given* is “information [the speaker] believes the listener already knows and accepts as true”, and *new* is “information [the speaker] believes the listener does not yet know”. In a similar line of thought, Chafe (1976) speaks about ‘information packaging’ and considers hypotheses about the receiver’s assumptions as crucial to discourse structure. These are hypotheses about the status of the referent of each linguistic expression, as represented in the mind of the receiver at the moment of utterance. Thus it is the way the information is transmitted that is crucial, rather than the lexical or propositional content of a sentence, around which grammar usually centers. Prince (1981:224) defines information structure (packaging of information) in the following way:

“The tailoring of an utterance by a sender to meet the particular assumed needs of the intended receiver. That is, information packaging in natural language reflects the sender’s hypotheses about the receiver’s assumptions and beliefs and strategies.”

‘Givenness’ has been attributed a formal status by Schwarzschild (1999) who claims that a given constituent is one which is entailed by the preceding discourse. This use of givenness is of course restricted to text-givenness, as opposed to context-givenness.

‘Topic’ or ‘given’ are often used interchangeably. The reason for this interchangeability is that many authors, like Kuno (1972), Lambrecht (1994) and others, comprehend topic as a phenomenon of consciousness and saliency, in addition to acknowledging the individual function of each concept as linguistic categories. For example, the following scale of ‘activation’ has been widely used in the literature: an active concept is given (it is then a topic), and an inactive one is new.

Activation: knowledge vs. consciousness (Lambrecht 1994, Chafe 1976):

- active concept: one that is currently lit up, a concept in a person’s focus of consciousness at a particular moment.
- semi-active (accessible) concept: one that is in a person’s peripheral consciousness, background consciousness
- inactive concept: one that is in a person’s long-term memory, neither focally nor peripherally active.

In the questionnaire, we regard a ‘topic’ as a referent which the remainder of the sentence is about (cf. Gundel 1988), possibly contrasting with other referents under dispute, and crucially followed by comment, typically containing a focus element. The topic has often been previously introduced into the discourse, but does not have to have been. We keep the notions of ‘topic’ and ‘given’ apart. We also use the notion of ‘accessibility’, although we are aware of the

difficulties attached to this notion when it comes to cultural peculiarities, since what is accessible or inferable in one culture may be inaccessible in another.

‘New’ may be understood as the complement of ‘given’ and ‘accessible’, whereas ‘focus’ is the complement of ‘background’ and ‘topic’ that of ‘comment’. These concepts are not excluding each other, because a given element may be focused, and a new element can appear as the topic of a sentence.

According to Rooth’s alternative semantics (1985, 1992), a focused constituent is expressed with a ‘focus semantic value’, which is an additional semantic value, $[Mary]^f$, besides the ordinary semantic value $[Mary]^o$. The alternatives to the focus play a central role in interpreting focus semantically. In the case of a contrast, as in (1b), it is the contrasting element *Anna* from the set of possible values of x in ‘ x likes Sue’.

- (1) a. $[Mary]_F$ likes Sue
 b. No, $[ANNA]_F$ likes Sue.

(Who likes Sue?: {Bill likes Sue, Mary likes Sue, Anna likes Sue})

The focus semantic value of a sentence is a set of alternatives from which the ordinary semantic value is drawn, or a set of propositions which potentially contrast with the ordinary semantic value. It is important to note that the ordinary semantic value is always an element of the focus semantic value.

Summing up, for the sake of the present questionnaire, it is important to distinguish between the status of referents in the mental states of interlocutors, which can be new (inactive at the point of their introduction into the discourse) or given (active), and the linguistic means which serve to distinguish between focused elements (designated expression in a set of alternatives), and backgrounded elements (like anaphoric or phonetically repeated expressions), as well as between topics (serving as the main referent for the remainder of the sentence) and their comment including focal information.

2 Grammatical Correlates of Information Structure

This section reviews the grammatical means for the expression of the main information structural concepts. These means are varied: they can involve the prosody, or even the segmental phonology, the morphology and the syntax. In the case of prosody, pitch accents are used to express focus and topic, especially in intonation languages; other changes in F_0 , such as boundary tones, register and tone scaling are widely used in different types of languages. In syntax, it is word order and changes in the grammatical functions of the arguments which are often used: cleft-sentences, topicalization and the like are syntactic strategies, motivated only by special information structural requirements. In morphology, particles for special kinds of focus are found in all languages. Some non-intonational languages identify narrow focus by means of special inflectional markers on the arguments themselves or on the verb. In some languages such particles have been shown to be the result of grammaticalization of auxiliaries.

The primary indicator of focus in intonation languages like English and German has repeatedly been claimed to be a falling accent, with topics being realized with a rising accent (Büring 1997, Steedman 2000, Jackendoff 1972, Selkirk 1995). A backgrounded element, that is an element which is neither a topic nor a focus, is generally deaccented (Schwarzschild 1999, Ladd 1980). Also crucial as a grammatical marker of newness vs. givenness is the use of articles: a new referent is introduced with an indefinite article and a given one with a definite article. The third factor which is often said to play a role for the marking of information structure in Germanic languages is word order. Several researchers find that the most common order is new after given (Hawkins 1994, Arnold et al. 2000, Clark & Haviland 1977).

Jacobs (2001) identifies German constructions which usually, and prototypically, express ‘topic-comment’ (left dislocation, hanging topic left dislocation, free topic, I-topicalization) and identifies grammatical devices widely used in expressing this distinction. In addition, he distinguishes properties of these prototypical topic-comment constructions: separation and predication are easier to classify as grammatical devices than addressation and frame-setting; a rising accent is an important component of a topic construction, as exemplified by the notion of ‘I-topicalization’, where ‘I’ stands for intonation. Buring (1997) goes a step further, and identifies a rising tone on the topic in German (see also Jackendoff 1972 and Steedman 2000 who also associate information structural elements with their prototypical intonation for English). Frey (2000), following Rizzi (1997) and many others, relates a topic with a syntactic position. For Frey, a topic is always located in a syntactically specific position that he calls ‘Topic-Phrase’ and which is situated above the I-Phrase. Buring (1997) and Krifka (1999) have a semantic approach to this notion: a topic is that referent which provides a partial and disputable answer to a question. Lambrecht (1994) insists on the referentiality of ‘topics’ and shows that discourse referents may be either entities or (less commonly) propositions.

3 Structure of the Questionnaire

The publication of the questionnaire¹ is divided into six chapters. After an introductory chapter, chapter 2 provides an instrument to deal with the grammar of languages to be investigated. A set of questions bearing on the phonological, morphological, syntactic and semantic structure have to be filled in by the

¹ There have been preliminary versions of this questionnaire: a first version in 2003, and a second one was released in June 2004. Both versions have circulated and served as basis for research in several languages.

researcher, who is in the ideal case a native speaker of the language. These questions lean on a long tradition of typological questionnaires such as Comrie & Smith (1977), and recent archives of typological features (Bickel & Nichols 2000, Brown et al. 2006, Corbett et al. 2006, König et al. 2006, to cite just a few). The aim of these questionnaires is to allow for cross-linguistic comparisons in terms of more or less standardized sets of grammatical properties.

Chapter 3 contains experimental tasks that are described in detail. Since they make up the main part of the research agenda, they are addressed further in the next section of this introduction. The manual contains descriptions of tests aimed at eliciting spontaneous sentences or short dialogues with specific information structural content. Different kinds of material accompany the experimental tasks: pictures, playing cards, and short films. The tasks have different forms and different functions. Some of them elicit just one sentence, while others trigger a whole dialogue.

Chapter 4 of the questionnaire contains translational tasks, i.e. sets of sentences to be translated both orally and in written form, completing the experimental tasks. The aim of this part is to provide a complete list of the different ways of expressing information structure, and in particular any strategies which may not have been used spontaneously by the informants while performing the experimental tasks. It thus provides a systematic control of the range of linguistic means used to express different information structural notions, and complements the results obtained from spontaneous production.

Elicitation through translation has largely been used in language typology, from the beginning of language comparative studies (see for instance the translation tasks of Bouquiaux & Thomas, 1987, for languages without literary tradition) up to recent typological studies (see Dahl 2000). Several viewpoints about the validity of data collected through translation in Newman & Ratcliff

(2001) highlight the potential priming effects of the contact language on the resulting data, as well as the qualitative differences between data elicited in this way and real spontaneous communication. Nevertheless, translation remains a valuable method of eliciting data for comparative purposes, and proposals have also been made about means to restrict the methodological disadvantages (especially with respect to priming effects, see Dahl 2000).

Chapter 5 is devoted to the interpretation of the data set collected through the QUIS. It provides some hints which grammatical forms can be expected to express certain information structural categories.

The questionnaire ends with chapter 6 containing information concerning the performance of the tasks in the field, and forms for documentation of field sessions (field session metadata; informant's agreement).

4 Experimental Tasks²

The experimental tasks listed in Chapter 3 of this manual are very much inspired by psycholinguistic and sociolinguistic research, as well as by tools for linguistic fieldwork such as those developed by the Cognitive Anthropology Research Group at the MPI in Nijmegen. The tasks use non-verbal stimuli for the collection of comparable data across languages.³ Of particular importance for our aim is the absence of priming as to which module(s) of grammar (prosody, word order, morphological markers) are to be used in a particular situation.

² A summary table of the experimental tasks is provided in Table 1 below.

³ Although not included in our questionnaire, we acknowledge the *pear stories* (Chafe 1980) that have been used for the study of information flow in narratives, as well as the *fish film* (Tomlin 1997) designed to investigate passive sentences.

The main types of tasks are:

- *Description of single situations.* This type of task is especially used for the elicitation of all new sentences through the description of pictures (see experimental task 10).
- *Description of sequences of situations.* A first picture, introducing a context situation, is presented to the informant, and then, in a second step, a picture showing the target situation. The informant's task is to give a short oral account of the sequence. In this setting, the discourse status of the referents in the target situation is manipulated through the choice of context situations (see experimental tasks 7, 3, 4, etc.). Similar manipulations are performed with short films instead of pictures (see experimental tasks 2, 23) and power point presentations (see experimental task 6).
- *Narration (of sequential events).* The informant narrates a story according to a picture series (in realis as well as in irrealis, see experimental tasks 1, 16, 19).
- *Picture discrimination game.* This is a collaborative task with two informants, one person in the role of the leader and a second person in the role of the matcher. The leader describes a situation and the matcher has to choose among alternative situations presented in different stimuli following the description of the leader (see experimental task 26).
- *Questions/answers.* The informant has to answer a question about a visual stimulus (picture or short film). Context conditions are established through different types of question: *wh*- questions, truth value questions, questions inducing correction, alternative questions, multiple constituent questions, etc. (see experimental tasks 1, 16, 17, 18, 16, 19, etc.).

-
- *Stimuli-matching games.* In an interactive game, two informants see slightly different stimuli and perform tasks targeting the differences. Experimental task 16 uses short films and picture series which differ in one or more crucial details. Experimental task 20 is a traditional map task.
 - *Instructing games.* In another type of interactive game, one informant (= leader) plays the role of an instructor, and another person (= matcher) performs a collaborative task. E.g., in experimental task 8, the leader describes a spatial configuration. The matcher is instructed to configure her/his cards according to the description.
 - *Role-playing games.* In this type of task, the informants are instructed to play the role of an individual in a story presented either through pictures (see experimental tasks 1, 9, 21, 19) or through films (see experimental tasks 16, 21) and to perform some conversational task.

The design of the experimental tasks follows current standards in the factorial organization of experimental items. Each experimental task is based on a number of *conditions*, which correspond to the discourse situations that are empirically compared. These conditions are implemented in an equal number of items, such as different pictures presenting different situations, in order to reduce the possibility of the resulting generalizations being influenced by situation-specific or stimulus-specific effects.

The experimental tasks of Chapter 3 are divided into four field sessions. In this way, each informant is confronted with each experimental item only once.

Table 1: Experimental tasks (by type of task and information structure category).

		<i>information status</i>	<i>focus</i>	<i>topic</i>
<i>descriptions</i>	<i>single situations</i>		10 Event Cards	
	<i>sequences</i>	1 Changes 2 Giving 3 Visibility 4 Locations 5 Sequences 6 Dynamic Localization	11 Anima 12 Contrast 13 Animal Game 14 Properties	22 Events in Places (C) 23 Path Descriptions 24 Groups 25 Connections 26 Indirect
<i>narration</i>			15 Eventives (C) 16 Tell a Story (A/B)	19 Fairy Tale (A/D)
<i>question & answer</i>		7 Birthday Party	15 Eventives (A/B/D) 16 Tell a Story (D) 17 Focus Cards 18 Who does what? 19 Fairy Tale (A/B/C)	22 Events in Places (A) 27 Surprises 28 Doing
<i>games</i>	<i>picture discrimination</i>			(26 Indirect)
	<i>stimuli-matching</i>		16 Tell a Story (C)	
	<i>instruction-giving</i>	8 Static Localization		
	<i>role-playing</i>	9 Guiding	15 Eventives (A) 19 Fairy Tale (B/C)20 Map Task 21 Drama	

5 Technicalities

- *Equipment.* All tasks are conducted orally using a good quality recorder (DAT, Minidisc, MP3).
- *Preparing the field sessions.* Questions or context sentences are provided online by the interviewer if she/he is a native speaker, or recorded before the performance of the session.
- *Duration of field sessions.* A field session takes about one hour. Obviously, the actual duration of the session depends on the individual performance and the situation (for instance, the need to translate online increases the time it takes to perform the tasks).
- *Documenting the session.* A form called “field session metadata” gives information about time, place, and informants (see Chapter 6).

6 Archiving

All collected data are gathered in a database. Part of the data are to be transcribed and annotated according to a separate annotation manual (<http://www.ling.uni-potsdam.de/~goetze/sfb/guidelines.html>). Annotated data is saved in XML format using the editor EXMARaLDA (see Schmidt 2004) and together with the audio files it is accessible in the database ANNIS (see <http://www.sfb632.uni-potsdam.de/annis/>; see Dipper et al. 2004). In ANNIS, a large (and expandable) number of annotation layers are available: information structure is the crucial part, but phonological features with special emphasis on intonation, morphological transcription, part-of-speech tagging, constituent structure, and semantic properties like animacy or quantification, are also annotated. Current typological conventions have been consulted, such as the guidelines for morphosyntactic glossing in Eurotyp (see König 1993) and LGR (see Bickel et. al. 2004) and current standards for the annotation of corpora such

as EAGLES. The results are available for the linguistic community under certain conditions of best practice, with citation and intellectual property being the main issues.

7 Acknowledgments

We have collaborated with a number of experts in the development of the various components of this manual. For the elaboration of the general questions (Chapter 2), we would like to mention Joanna Blaszczak and Stefan Hinterwimmer, and for the development of the experimental tasks (Chapter 3), Cornelia Endriss, Robin Hörnig, and Thomas Weskott. Special thanks are due to Rainer Dietrich for his suggestions concerning the design of production experiments for the investigation of information structure.

The following associate researchers to our project have contributed to the development of the questionnaire with helpful comments and practical issues from the field as well as methodological problems in specific cultures and object languages: Alain Thériault (Can. French), Bei Wang (Chinese), Caroline Menezes (Konkani), J. Diego Quesada (Teribe), Yukiko Morimoto (Japanese), Kriszian Tronka (Hungarian), Ruth Singer (Mawng), Hanneke Van Hoof (Dutch), Wolfgang Sperlich (Niue), Teng Shi Chi (Prinmi), Rusudan Asatiani (Georgian), Yannis Kostopoulos, Thanasis Georgakopoulos, and George Markopoulos (Greek), Elizabeth Medvedovsky (Amer. English).

Following artists have contributed to the development of stimuli for QUIS: Pär Wrestling (pictures in task 1: items 1-2; task 9: item 1; task 19: item 1; task 28), Paul Starzmann (pictures in task 9: items 2-4; task 16: items 5-8), Annelies Schwarz (pictures in task 15: items 3-4; task 19: item 2), Stephane Leonard & Claudio Pfeiffer (movie ‘The broken vase’ in task 21: item 2). All other pictures have been made with the software *Poser* (Version 5.0).