Multiple languages and multiple methods: Qualitative and quantitative ways of tapping into the multilingual repertoire

Raphael Berthele

1. Introduction

In this contribution, I will discuss different methods that can be used to empirically investigate the dynamics that are at work in multilingual language usage. More specifically, using examples from research on comprehension in foreign or unknown languages, I will consider the possibility of combining different methodological approaches. Methodological discussions tend to be rather uninteresting when they are tantamount to stereotypical attacks at straw men on the far and near side of the qualitative-quantitative divide (Bergman 2008b). The recurring invocations to overcome this divide by applying multi-methods approaches and so-called ‚triangulation' are certainly well-intentioned but often problematic due to insufficient conceptual clarity about what type of evidence lends itself to triangulation and what its epistemological status in each particular case could be (see below, Section 2.3).

The very general remarks on methodological choices in this and the following section can by no means replace the extensive literature on (applied) linguistics methods (Seliger and Shohamy 1990; Nunan 2008) and methods in social sciences in general (Atteslander, Bender, Grabow, and Zipp 1991; Bortz and Döring 2003). The goal of the following paragraphs is to contextualize and to motivate the methodological considerations that underlie the studies that will be reported in the main part of this contribution.

This contribution is organized as follows: Section 2 provides some general considerations on methodological choices and the combination of methods in general. One of the most important points of Section 2 is that methods vary in their potential to limit the type data that can be gathered (control dimension) and in their degree of intervention or invasiveness into the language users' universes. These and other fundamental methodological notions are then illustrated in Section 3 by means of two case studies. The case studies show how different elicitation methods and data types yield different types of results that, in some cases, can be combined in the sense of mixed-method approaches. The first case study taps into the role of grammatical structures in comprehension of German as a foreign language, and the second study investigates the capacity of multilinguals to infer the meaning of cognate words in foreign languages. In the final section, concluding remarks sum up the main points of the paper by shedding light on critical aspects of both selective and non-selective approaches to language usage realities in multilingualism research.

2. Methodological choices as problem-solving procedures

2.1. Minimal prerequisites for scientific endeavors

In the remainder of this chapter, the term research refers to linguistic endeavors that involve at least three elements: A problem, data, and interpretation. Data without a research question (=problem) do not represent research, and neither do purely theoretical constructs without data (in the broadest sense of the word). In standard falsificationist approaches, the problem precedes the data, but some researchers also allow for the reverse logic, e.g. within the grounded theory approach (Glaser and Strauss 1967), where theories are supposed to emerge from the data. For the sake of maximal inclusiveness these approaches also fall under my definition of research, although I am personally very skeptical that it is ever possible to collect and analyze data without any (implicit) theories or assumptions as points of departure.

The data can be a corpus of text, experimentally collected responses, results of observation or of introspection, as has been the practice in early generativist linguistics. The term *problem* is intentionally used here and should not be confounded with the term *hypothesis*. Whereas the latter entails a particular type of research (hypothesis testing), the former is more general and deliberately involves all kinds of subject-matter related interests and goals a researcher can have. Science, in this view, is a continuous and principled attempt to solve problems (cf. Popper 2004).

A scientific problem is embedded into a theory in the broader sense of the word, i.e. a more or less consistent model involving assumptions about concepts and causal relationships between them. The goal of science is twofold: problem-solving and theory development. In the epistemological perspective taken in the present contribution, theories cannot be proven directly by research; they rather develop in a process of rejection, refinement, and revision on the basis of evidence collected by researchers. I expect any research plan – be it a radically qualitative or a radically quantitative and falsificationist one – to bear the potential of yielding results that force the assumptions underlying the research plan to be revised. Causal theories (e.g. "individual multilingualism fosters creativity"), for instance, require experimental methods that bear the potential of failing to support the theory. Ethnographic or other qualitative data cannot be used to 'prove' or falsify such hypotheses, due to the lack of control of relevant factors and the high risk of confirmation bias in the selection of informants, settings and data items. On the other hand, if researchers aim at understanding or describing creative interaction among multilinguals in institutional settings, qualitative methods are a possible and sensible approach. They bear the potential of providing new evidence that ultimately leads to new and better theories of the nature of multilingual creativity.

2.2. Control over data and degree of intervention

Any research plan can be specified along at least two dimensions: intervention and selection (cf. van Lier 1988). Firstly, researchers have to decide whether they are looking for very specific kinds of data (a particular dis-course marker, plural morphology, conditionals, etc.) or not. Depending on this choice, methods that provide sufficient amounts of data in the required quality will be chosen: searches across corpora, questionnaires, translation data for controlled studies, unfocused corpus data, ethnographic and interview data for less controlled research plans. The selection dimension is a gradual one, i.e. one can easily imagine data collection procedures that aim for a certain amount of control without going to extreme lengths in order to elicit, say, a high density of conditional verb forms by using a very focused elicitation procedure.

The second dimension is represented by the degree of intervention: On the high degree of intervention end of the scale there is a fieldworker who administers often rather unusual tasks in laboratory conditions, i.e. in conditions that are markedly different from 'normal' contexts of language usage. On the other end, there is little to no intervention from the researcher, either by working on data sets that have been collected for purposes totally outside the specific research project (e.g. corpora of newspaper articles), or by collecting observational data in ethnographic projects. Again, there are degrees of intervention, the psycholinguistic experiment being an extreme endpoint on a scale.
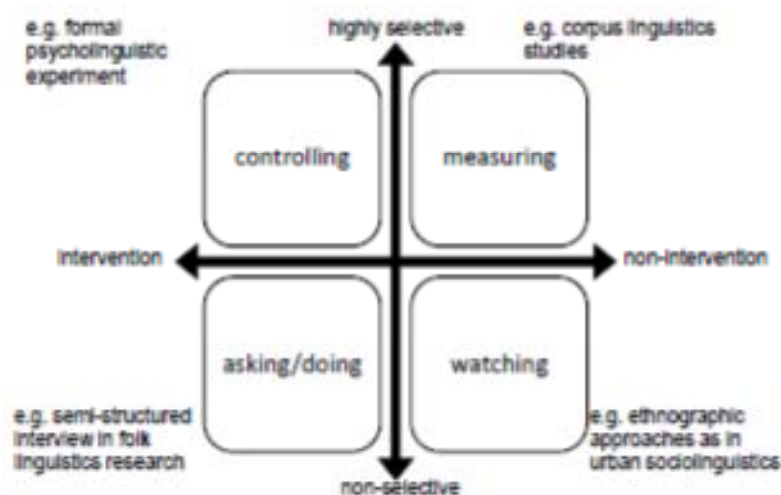

Figure 1. Two dimensions and four spaces (based on Nunan 2008: 8)

Sometimes, research from the watching space in Figure 1 is associated with a 'constructivist' vision of reality whereas research from the controlling space is deemed (or condemned) to be of the 'positivist' type. This association of visions of reality and research methods is too simple: It is perfectly possible to carry out ethnographically inspired research while having a rather positivist stance. Quantitative research, on the other hand, does not necessarily presuppose a realist or objectivist point of view (as opposed to nominalism or constructivism): There is no fundamental incompatibility between a constructivist take on reality and quantifications. Just as in qualitative research, the researcher can presuppose the categories or entities that he/she operationalizes to be socially and cognitively constructed. One of the fundamental categories used in linguistics, that of a *language*, is a good case in point for the construct character of the object of inquiry, since it is notoriously unclear where the boundaries of a particular language and/or dialect should be drawn.

Although most research predominantly uses one type of methodology, in many cases methods are combined. As in the research project discussed below, there are reasons to change perspectives and to use methods and data pertaining to two or even more spaces in order to find answers to the respective problems. Such combinations are often referred to as 'triangulations'.

2.3. Triangulation

In trigonometry triangulation allows determining the location of a target point by measuring angles to it using a baseline defined by two points. There is a certain tradition of mapping triangulation metaphorically onto the domain of methodology in the social sciences (Bryman 1992; Denzin 2005; Bergman 2008a; Hammersley 2008). Interestingly, taking the notion of triangulation literally implies that the determination of the target point is not possible without a basis of at least two different known points of view. A direct mapping of the trigonometry meaning onto the scientific target domain would imply that no scientific result can be obtained based on one point of view, which would certainly be an unusual and radical claim. Moreover, many usages of triangulation involve combination of data that are totally different in nature (e.g. experimental and ethnographic data). This is fundamentally different from trigonometry, where the two starting points are necessarily of the same kind.

Triangulation in social research, following Hammersley (2008), can have the four functions listed below. These functions are not all mutually exclusive (see also Blaikie 1991; Flick 2008 for critical discussions of the concept of triangulation):

1) validity checking: by using other data sources (e.g. by combining different quantitative measures, or by combining qualitative and quantitative methods)

2) indefinite triangulation: make visible how accounts are shaped by different purposes/perspectives of social actors

3) seeking complementary information: (probably the most common usage); can lead to correction of the first interpretation and is thus not incompatible with the validity checking function

4) epistemological enrichment: transgress the limitations of particular methods by combining several approaches; encourage dialogue between paradigms

The examples of triangulation given below will primarily serve the third function: The goal is to seek complementary information since the first research paradigm chosen has given rise to new questions and problems that could not be resolved based on, e.g., quantitative data only.

2.4. Focus on experimental and other psychometric methods

Historically, research on bi- and multilingualism starts with the comparison of linguistic systems which, in a second step, allows predictions about do-mains of potential interferences in bi- or multilingual individuals (Weinreich 1953; Lado 1957; Ringbom 1990). A prerequisite of this approach is knowledge of the 'systems' that enter in contact and the empirical basis often is corpus data from bilinguals, collected in more or less natural environments or in language learning contexts (cf. Lado's contrastive analysis hypothesis). Generally, this research ranks low on the intervention parameter but high on the parameter of selection, i.e. since looking at all aspects of linguistic structure at the same time is hardly possible, researchers focus on particular partitions of linguistic structure. Only in more recent times have experimental data entered the field of multilingualism research (Gullberg, Indefrey and Muysken 2009). The term experiment can be used in a narrow and in a wider sense:

A) Experiment in the narrower sense ('true experiment'): Control for all relevant variables (ideally), laboratory conditions, pre- and posttests, experimental and control groups, random assignment of participants

B) Experiment in the broader sense: Control for a maximum of relevant variables, field conditions, pre- and posttesting and experimental/control groups can be replaced by post-hoc grouping of participants according to selected independent variables, no random assignment of participants

There are different terms in the methodological literature for research that can be attributed to the second category above. All of these types of research are located somewhere in the controlling space in Figure 1, with slightly varying degrees of

selection and intervention. Field experiments rank somewhat lower on the intervention scale than true experiments since they do not require laboratory conditions which by nature are highly invasive. The field and quasi-experiments with multilingual subjects discussed in Section 3 do without randomization of subjects and thus cut back on the intervention dimension as well, but they can still target quite selective data types, e.g. cognate recognition as investigated in Berthele and Lambelet (2009), or transfer of syntactic schemata from one language into the other as investigated by Peyer, Kaiser, and Berthele (2010). In other cases, less selective data types can be elicited, e.g. when measuring global text comprehension.

As in many other disciplines, researchers in multilingualism studies are interested in causal relationships between variables. Does bilingualism help in learning a third or additional language? To which extent do structures of a multilingual's first language causally determine the dynamics of second or third language learning? Does bilingualism foster intelligence? The method that best licenses claims about causality is a true experiment (Waldmann 2002). Unfortunately, true experiments are often impossible to carry out. Controlling for bilingualism and all other relevant factors in an experimental paradigm would require a representative sample with random assignment of participants to the experimental (bilingual) and control (monolingual) group and then training the bilingual group in a second language, which even according to the most liberal definitions of bilingualism would take several years. Such designs are unrealistic, which is why researchers are forced to draw on other methods, trying to control as many factors as possible, but easing methodological restrictions such as randomassignment or experimental treatment and pre- and post-testing. Quasiexperimental and ex-post-facto research thus is often the best multilingualism researchers can do, if they go for the controlling space at all. The price to pay is that causal relationships between variables can hardly ever be tested.

3. Empirical investigations on the multilingual repertoire at work

In this section, evidence from two research programmes on receptive proficiency in multilinguals will be presented and discussed. Since the emphasis for the present purpose lies on the methodological choices, the description of the samples and procedures is not comprehensive but rather selective.

3.1. Example 1: Investigating the role of grammar in comprehension of German as a foreign language

The main goal of the first project to be discussed here is the investigation of the role of grammar in understanding German as a foreign language (see Peyer, Kaiser, and Berthele 2010; Kaiser and Peyer 2011 for details). Although there is a considerable number of contrastive grammatical analyses of German and some other Western

European languages, there has been only very little empirical work on their actual measurable influence on exolingual comprehension. The way we wanted to empirically investigate the role of contrastive features of German grammar was to proceed in a multimethods approach combining non-selective observation data with testing data of a more selective type, with each stage in the research plan pursuing particular epistemological goals (see Table 1 for an overview).

Table 1. 3 phases. The selection and intervention scales are roughly divided in a high, middle and low segment (+, ±, –).

Phase	(1) Qualitative pilot phase	(2) Quantitative main phase	(3) Qualitative feedback phase
Main goal	Identify potential problem zones (grammar)	Test empirical difficulty of previously identified constructions	Observe how learners try to understand items that yielded surprising quantitative results
Method	Elicitation of verbal protocols of translation of authentic German texts into L1	Assessing and testing: Proficiency tests, comprehension task with discrete point test items	Verbal protocols of group task similar to (1), translation of selected passages from the reading texts in (2) into L1
Data type	–selective; ±intervention	+selective; +intervention	±selective; ±intervention
Sampling	Students of German at Italian and French universities, L1 French, Italian.	Mostly students of German at Italian and French universities, L1 French, Italian.	Students of German at Italian and French universities, L1 French, Italian.
Analysis	Mainly bottom-up search in the data for grammatical problems	Hypothesis testing of empirical difficulty of target structures vs. alternative structures	Look for new explanations of surprising patterns emerging from the quantitative data

To illustrate the type of evidence collected in each phase of research, I will briefly discuss three examples, all of which concern the same grammatical characteristic of German grammar, viz.the possibility to have OVS syntax.

The first example is a token from phase (1), produced by a Francophone law school student, who is at an advanced intermediate level in German and enrolled in a bilingual MA programme at the bilingual University of Fribourg. We selected newspaper articles on law issues from the *Neue Zürcher Zeitung*, and asked the students in a one-to-one setting to read these texts and to translate them into their native language. The goal was to observe where the multilinguals can provide smooth and adequate translations and where the informants hesitate, make mistakes or are completely lost. We always provided help with respect to vocabulary, either in the

form of lists or in the form of online oral translation of particular items the informants were struggling with.

(1) Target item from a newspaper article: Dem Bundesgericht erscheint die Einschränkung

the-DAT federal_court appears the limitation der Urlaubsgestaltung keineswegs als verfassungswidrig. the-GEN vacation arrangement in no way as counter the constitution 'The federal court regards the limitation of the vacation arrangement in no way as counter the constitution.'

Participant: Noëlie, Law Student at University of Fribourg, L1 French

A possible explanation of the problem in [1] and [2] above is that Noëlie's first assumption is that the first NP constituent in the target item functions as the clause's subject, which is not the case in this particular example. This assumption could be based on the strong cue to subjecthood represented by an NP in clause initial position in French, whereas this cue in German is less reliable. It could further be hypothesized that it is only after the intervention of the fieldworker that the participant notices the dative morphology (<Dem>) in the definite article and revises the argument structure in her interpretation of the clause. It could be that the informant applies some sort of a 'natural' default parsing strategy where the agent and thus usually the subject is expected to be in first position. Or, from a multilingualism point of view, it makes sense to assume that the subject's L1 with its deeply entrenched syntactic SVO schema interferes with the parsing of the German clause. However, we need to keep in mind that the method chosen for this pilot study does not control for important factors: We don't know for sure that the sentence would have been translated more easily if the topological order of the L1 had been respected (SVO), since other factors, such as the general amount of semantic complexity of the clause could interfere. And we don't

know whether the token is an accidental drop in receptive performance of this particular informant or whether she has problems in general with OVS structures, or with dative morphology. Moreover, we don't know whether this hypothetical transfer of L1 syntax is only a phenomenon we encounter on low or intermediate levels of German proficiency. To sum up, although the SVO-transfer theory is intuitively plausible and seems consistent with the qualitative data gathered in the first step of the project, there are problems of reliability and validity that do not allow hasty generalizations. This is why we need quantitative methods as well.

Although the method used in this phase is clearly qualitative, some quantifications were done based on the verbal protocols: If several informants on several occasions showed problems of the type illustrated by Ex-ample (1) above, we decided that the structural properties shared by the different items were good candidates for a structure to be tested on a larger scale in phase (2). In this regard it is important to note, however, that the type of qualitative evidence, although it is in some sense richer than quantitative operationalizations, does not 'represent' or 'mirror' reality, but is just as well based on interpretative operations.

The next phase in our research project was to select a number of promising structures from phase (1) and to construct a new instrument that allows for controlled quantitative data elicitation in the sense of the upper left space in Figure 1. As described in Peyer, Kaiser, and Berthele (2010), the goal here was to control for certain factors influencing reading in German as a foreign language and to shed light on the role grammatical aspects play in comprehension. The most controlled fashion of doing this would have been an experimental setting involving e.g. target sentences containing particular structures (experimental condition) or not (control condition) presented on a screen while measuring reading and other response times as well as the comprehension of the sentence. This kind of setup allows for relatively high degrees of construct and internal validity, and, depending on the sample, also of external validity. However, and this is the great disadvantage of highly controlled studies, the ecological validity, i.e. the degree to which reading and comprehending isolated target items in a foreign language stand for the process of reading in more or less natural situations, is supposedly low.

Here, methods that involve observation (of reading processes, e.g. via verbal protocols) and maybe asking (about strategy usage) have much more to offer. The disadvantage of such methods, on the other hand, is that they do not provide the highly selective type of data we were aiming for. The relatively specific question of the role particular features of grammar play in comprehension of German as a foreign language could probably not be answered based on data collected in a non-selective paradigm. The solution was a compromise: We constructed pseudo-authentic texts in

the style of encyclopedia articles about imaginary animals, which allowed us to control for knowledge about the world while keeping the text type and the reading situation relatively natural. By giving written (interlinear) translations of content words in the texts we aimed at controlling for lexical knowledge, thus trying to isolate the grammatical component of potential comprehension problems. This methodological choice, however, deliberately dissociates two things that are never actually separated in 'real life', at least not if the theoretical presuppositions involve a monotonous construal of a grammar-lexis continuum. Arguably, a generativist stance might be more sympathetic to the division between the two 'modules'.

We worked with parallel versions of texts, which were propositionally identical but contained different grammatical structures. The structures that varied systematically were those deemed to be relevant based on the first phases of the research project (Table 1). For each sentence or clause containing the target structure (e.g. OVS) there was an alternative grammatical structure which expressed the same proposition in the parallel version of the text (e.g. SVO; cf. Example 2).

(2) Example of target and alternative structure

OVS	SVO
Einen Boren fressen Flundodile gerne.	Flundodile fressen gerne einen Boren.
A:ACC boren eat flundodiles willingly	Flundodiles eat willingly a:ACC Boren.

The participants in the experiment, a total of 506 Francophone or Italianspeaking university students with varying levels of proficiency in German, had to respond to discrete-point test items tapping into the exact comprehension of the target items. In the case of Example (2) the comprehension test was simply an item asking to list the food the Humpfhorn eats.

The OVS structure was operationalized by items such as Example (2) above. On the whole, this structural characteristic of German turned out to be empirically difficult as compared to equivalent structures of the SVO type (see Figure 2). The difference in empirical difficulty between the two structures is statistically significant overall as well as for all levels from B1 on upwards (Kaiser and Peyer 2011: 194). The impression we had gained from the analysis of the qualitative data thus led to a hypothesis that survived inferential statistical testing in the controlling space of Figure 1.
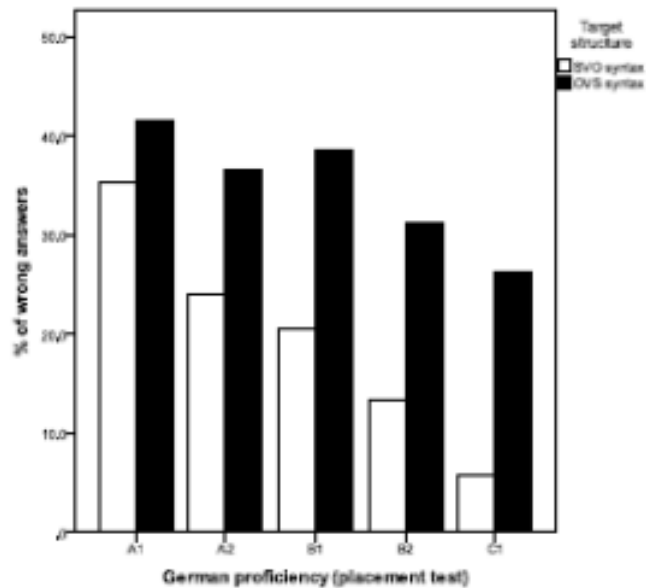
Figure 2. Empirical difficulty of items containing OVS vs. SVO structures

As far as the question of external validity is concerned, we deem the result to be generalizable to reading situations involving rather detailed, informationally dense, technical texts read by native speakers of either Italian or French with high literacy skills.

Whereas the global results for the OVS structure suggest an increased difficulty of items following this noncanonical (from the point of view of a native speaker of an SVO language) pattern, the analysis of individual target items in the stimulus texts sometimes revealed surprising results. As a last example, the item in (3) will be discussed.

(3) Example of target and alternative structure

OVS	SVO
Dem Humpfhorn dienen aber in seltenen Fällen auch Mödele und Blusten als Nahrung	Aber auch Mödele und Blusten dienen dem Humpfhorn in seltenen Fällen als Nahrung
The DAT Humpfhorn serve however in rare cases also mödele and blusten as food	However, also mödele and blusten serve the DAT fhundodile in rare cases as food

This item produced a mean error rate of 33.3% (OVS) vs. 26.2% (SVO). This difference is statistically not significant. Based on this item only we were thus unable to confirm the hypothesis regarding OVS difficulty. Such results, which go against our expectations, emerged also for some of the other target structures. As is often the case, although some answers can be given on the basis of the hypothesis testing (falsificationist) paradigm, more new questions arise from the patterns found in the data. As shown in Table 1 above, the last stage in the research project was to return to more qualitative, observational data in order to generate new hypotheses about the particular items that are not 'well behaved' in the sense of the general contrastive approach that had fostered the initial hypotheses.

Example 4 is a transcript from this third phase. Adult learners of German as a foreign language were presented with our target texts again, and the task here was to work in dyads on the comprehension of these texts, while thinking aloud and translating orally into the participants' L1.

(4) Verbal protocol of translation. Target item cf. (3) above.

[1]

	0 [00:00.0]	1 [00:02.0]		2 [00:04.0] [00:06.0]	4 [00:08.0]
A [v]	alors euh dans certains cas le humpfhorn	euh	se sert euhh		
B [v]					mais dans
A [v]	well mmm in certain cases the humpfhornmmh	uses mmhh			
B [v]					but in certain

[2]

	5 [00:10.0]	6 [00:12.0]	7 [00:14.0] [00:16.0]
A [v]	dans de rares cas le hum le humblom le humpfhorn (2.0)		
B [v]	certains cas		sert
A [v]	in rare cases	the humblom the humpfhorn	
B [v]	cases		serves

[3]

	9 [00:18.0] 10 [00:22.0]	11 [00:24.0]
A [v]	(4.0) Ah 'Dem Humpfhorn dienen aber in seltenen Fällen such' (2.0)	
A [v]	Ah, "But the:DAT Humpfhorn serve in rare cases also"	

[4]

	12 [00:26.0]	13 [00:28.0]	14 [00:30.0] 15 [00:35.4]
B [v]			C'est bizarre. Mais
C [v]	Il sert de nourriture aux mörkele et aux blusten mi (5.0)		
B [v]			That's strange. But I
C [v]	It serves as food to the mörkele and to the blusten		

[5]

[5]

		16 [00:38.0] 17 [00:40.0] [00:53.0]
B [v]	je comprends pas pourquoi y a du (-) du du datif(-)	ah si c'est euhh mais
B [v]	don't understand why there is	a dative	ah, if it's mmmh but in

[6]

	19 [00:56.1]	20 [00:57.7]	21 [01:02.9]
B [v]	dans dans de rares cas les mörkele et les blusten servente de (.) nourriture au		
C [v]		(.)les mörkele	nourriture (.) au
B [v]	rare cases the mörkele and the blusten serve as food to the		
C [v]		the mörkele	food to the

[7]

	22 [01:05.4]	23 [01:07.1]
B [v]	euhh humpfhorn	
C [v]	humpfhorn	
Fieldworker [v]		mhm (.) genau
B [v]	humpfhorn	
C [v]	humpfhorn	
Fieldworker [v]		mhm exactly

As can be seen from Example 4, at the very beginning of the translation attempt (Segment [1]) the informants generate a mental model of the text that is propositionally close to the meaning of the text. This is remarkable since they are far from parsing the sentence appropriately: The dative object is hypothetically analysed as the subject and the verb (*dienen,* 'to serve') is translated by *se servir*, which would have an argument structure that is equivalent to *use*. There seems to be, at least at the very beginning of the comprehension process, a conspiracy of shallow parsing of the constituent structure with a near mistranslation of the lexical verb which leads to a propositionally correct translation (*Humpfhorn eats mörkele and blusten*). As soon as the dative morphology is discovered, things get complicated and temporarily messed up. It is not until in the second to last segment [6], after rather long reflection and discussion of the sentence, that a mental model emerges that is not only propositionally correct, but that is also yielded by a correct analysis of both lexical content and argument structure. This analysis allows shedding light on the surprisingly high percentage of correct comprehension of this rather difficult item: If especially low proficiency learners of German only run a superficial analysis of the passage based on some conceptual content conveyed by lexical items and frame knowledge of eating (requiring an agentive eater and an eatee), the analysis can be propositionally correct despite considerable linguistic ignorance.

3.2. Example 2: Interlingual inferencing of cognate words

The last example to be briefly discussed here stems from a series of investigations into the way multilingual individuals infer the meaning of cognate words. Again, the focus lies on comprehension of a non-native language, but this time the target language is an unknown one, even though genealogically related to the multilingual subjects' previously acquired languages. Moreover, the target items are not sentences or texts, but only (cognate) words. As in Section 3.1, I will show how evidence from different methodological spaces (cf. Figure 1) can be combined and integrated to enrich the global understanding of the underlying processes.

The stimuli are presented either with or without context, and the target items are words in languages that the participants have not learnt, but that are genealogically related to languages they master. The broader context of these studies is the interest in intercomprehension or semi-communication (Braunmüller and Zeevaert 2001; Hufeisen and Marx 2007), and more specifically, in ways of increasing the usage value of 'smaller', lesser used languages via the rapid development of receptive competences. As several studies on different target languages have shown, there is weak but statistically meaningful correlation between the number of languages a participant speaks and the general ability to infer the target items. In particular one study (reported in Berthele 2011) provided evidence for positive correlation of this inferencing capability with the age factor (the older the better) as well as with modern

language learning aptitude. For this present methodological discussion one particular aspect will be in the focus: What are the characteristics of cognate words that are generally well inferred by multilinguals, and what are the characteristics of words that turn out to be impossible to infer? Based on ideas from research on third language acquisition and transfer (Odlin 1989; Ringbom 1990), the hypothesis was that the best predictor for the empirical difficulty of items should be the linguistic distance between the target cognate and the transfer base in the multilingual lexicon of the inferring individual. As illustrated in Figure 3, cognate words can be very similar or quite different with respect to their targets, if measured using string similarity algorithms such as the Levenshtein distances (cf. Heeringa, Kleiweg, Gooskens, and Nerbonne 2006). At least from a psycholinguistic point of view it seems reasonable to construe the category of cognate as a radial category with fuzzy boundaries rather than a clear-cut category based on genealogical relations across languages.
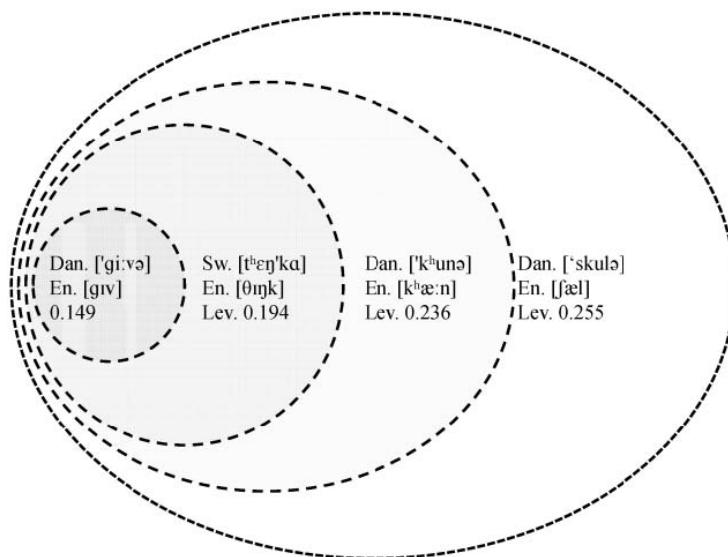

Dan. [ˈɡiːvə]	Sw. [tʰɛŋ'kɑ]	Dan. ['kʰunə]	Dan. [ˈskulə]
En. [ɡɪv]	En. [θɪŋk]	En. [kʰæːn]	En. [ʃæl]
0.149	Lev. 0.194	Lev. 0.236	Lev. 0.255

Figure 3. Danish and Swedish Cognates and feature-based Levenshtein distances to English

For the sake of brevity I will only give a short summary of the results of the experimental (in the wider sense) data on listening comprehension of cognates: The quantitative analysis of the empirical difficulty and the linguistic distance as measured by the phonologically weighted Levenshtein distances revealed that items that are beyond a particular threshold (cf. Figure 4, threshold around 0.22) are hardly ever correctly identified by the participants, but that for items below this threshold the correlation is weak.
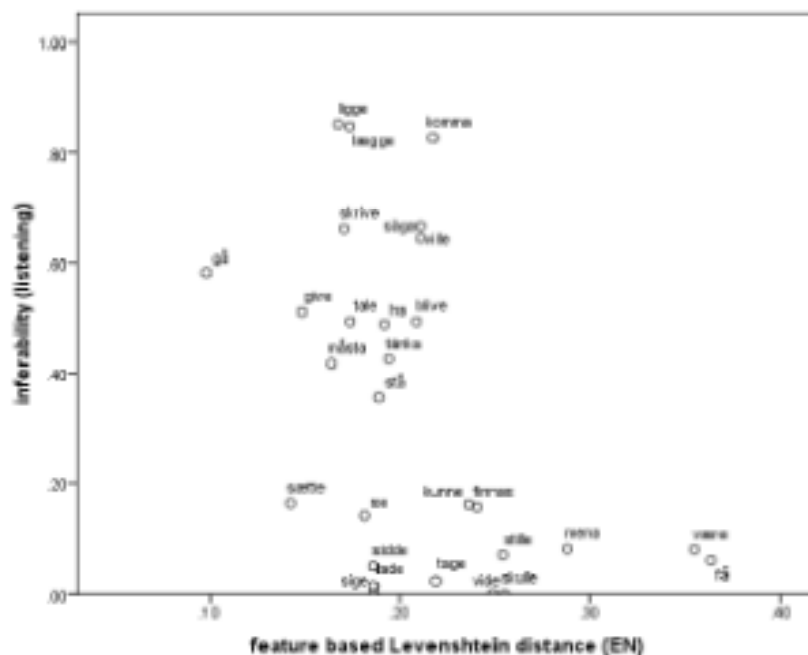
Figure 4. Empirical difficulty (inferability) of items and feature-based Levenshtein distances to English cognates

A subsequent quantitative analysis aimed at the detailed investigation of the impact of particular phonological contrasts between target words and potential transfer bases. The results suggest that as soon as consonants are different (with respect to place or manner of articulation), interlingual inferencing becomes very difficult. Phonological differences in vowels, on the other hand, do not seem to be a problem, in many cases a difference even coincides with better interlingual inferencing (cf. Berthele 2011 for a detailed discussion of these analyses). However, we cannot be entirely sure whether the patterns are internally valid, i.e. whether the items that could not be inferred with success by a majority of the informants remained opaque solely due to consonant differences (see also Beijering, Gooskens, and Heeringa 2008). One way of cross-validating this 'consonant theory' of cognate recognition was again a methodological change of spaces (cf. Figure 1): Moving from the controlling space of the quasi-experimental design to the asking-space of a thinking-aloud task with the same target items seemed to be a way of investigating the question whether multilinguals are indeed relying more on consonants than on vowels in cognate recognition. Below, five examples of verbal protocols (simultaneous thinking about the cognates) are given, from two different participants.

With respect to the 'consonant theory of cognate recognition' that emerged from the quantitative analyses, the examples above (and there are many more in the data) seem

to show that, indeed, the two participants search lexical entries in their multilingual repertoires for matching consonant skeletons. The item *skulle* (that has not been inferred correctly by anybody in the sample) is particularly revealing, since the participants associate all kinds of English and other words that share either the /sk/-/l:/ pattern (in one case with the addition of an /r/: *scroll*) or at least a /l:/ in second consonant position. Whereas there is no way of tracing the searches the participants carried out in the quantitative analyses of the paper-and-pencil variant of the task, the thinking-aloud task allows to tap into these interlingual pro-cesses. On the other hand, since the thinking aloud data are relatively nonselective in nature, quantifications are difficult and inferential statistics are impossible and/or make little sense. The triangulation (in sense 3 of the list in Section 2.3) of the quasi-experimental data from the controlling space and the verbal protocol data from the asking space seems to produce converging evidence in support of the consonant theory of cognate recognition, since many of the verbal protocols show how informants vary systematically the vowels while keeping the consonants fixed.

Since most of the data discussed here stem from rather highly educated participants (mostly university students), the external validity of the results remains limited to populations with similar educational backgrounds.

4. Conclusions

Research on multilingualism obviously obeys the same methodological constraints governing most other empirical disciplines. Science-internally it is most important that researchers be aware of the underlying epistemological stances of these activities. In my view it is too simple to associate qualitative methods with constructivism and quantitative methods with objectivism, as Nunan (2008: 4) seems to suggest: Bad ethnography is not subjective, but intransparent and perpetuates preconceived wisdom, bad psychometrics 'proves' by showing off statistical pomp ill operationalized constructs and unwarranted claims on causal relations between them. As a believing and practicing constructivist I would like to argue that both research belonging to the controlling space and to the watching space (e.g. psychometry vs. ethnography) construct their field/reality based on apriori questions/categories. Triangulations in the third sense thus are a delicate affair, since different elicitation methods lead to different construals of reality. For the examples from my group's research discussed above, one can thus object that the verbal protocols represent a type of data that is not directly comparable to the quantitatively gathered responses to the target items. I have tried to attenuate this problem by keeping the target items (grammatical target structures, cognate words) and the goal of the tasks identical (comprehension) by only varying the nature of the response data and the constraints on the context of the task. But I would not deny that there is what I suggest to call a "constructivist threat" to triangulation approaches that needs to be taken seriously. More generally, it seems that

hasty attributions of realism and nominalism to quantitative and qualitative approaches respectively are as inappropriate as the celebration of triangulation and multi-methods approaches as universal cure. Triangulation is not per se better than a single-method approach, and moreover it is important to distinguish the fundamentally different types of triangulation listed in Section 2.3.

By no means have I wanted to overrate the two examples of my own research presented in this contribution as particularly outstanding tokens of methodological excellence. There are obvious shortcomings of the methodlological choices, e.g. the second stage of the first project presented (Table 1) is only half-heartedly located in the controlling space, which was the price we paid for a minimum of ecological validity of the reading task. More rigorous control (i.e. even more selective data and more intervention) would be an important complement to our study. Other shortcomings of all the analyses presented here can easily be identified. The studies were thus merely intended as illustrations of two points: Firstly to show the particular constraints imposed by each methodological choice, and secondly to illustrate possibilities of moving across the different methodological spaces in Figure 1.

In addition to the scientifically relevant questions of the relation between methods and potential realities that are more or less independent of the beholder, multilingualism researchers need to pay particular attention to questions of validity: In modern Western societies issues that are at the core of our field are also in the focus of educational and migrational policies. We therefore need to be particularly aware of the degree of generalizability of the insights we gain from our data. Researchers, particularly if they are carrying out mission oriented research payed for by actors from the education policy domain, need to state clearly the threats to internal and construct validity that any operationalization bears, and they need to be very clear about the limits of generalizations inherent to their research paradigms: Qualitative analyses, despite their potential to provide "thick descriptions" of language usages and their contextual embeddings, must not be used as bases for generalizations unless there is substantial converging evidence from other studies applying other methods. Quantitative studies that do not involve adequate sampling techniques and control for the most important factors influencing linguistic competence (and there are many such potential factors) only license very limited external generalizations. Biased sampling practices in psychology or psycholinguistics (e.g. doing experimental research based exclusively on psychology students taking ECTS points for their participation) pose serious threats to the external validity of the research results.

Although these caveats are far from being new, we can frequently observe that language policy is based on insufficient scientific evidence or on abusive misinterpretation of scientific results. The misguided usages of the PISA survey data

in educational policy debates are only one very prominent example: A cross-sectional monitoring study that compares the efficiency of educational systems has been used as 'scientific proof' of all kinds of causal models that, from an epistemological point of view, could only be investigated via longuitudinal or experimental designs.

Methodological questions therefore are crucial especially for research in applied linguistics: Contrary to the somewhat less noble connotation of applied (as opposed to fundamental) research, it is precisely the applied linguists who need to be particularly aware of the epistemological constraints governing scientific activities and of the high methodological standards that are needed when social, educational and political issues are at stake. This is not to say that it is easy to be a structural linguist analyzing Icelandic passives, but it is at least as complicated and challenging to make scientifically valid statements about multiple first language acquisition in Swiss German kindergartens, let alone about which language policies in linguistically diverse contexts lead to more educational success for members of (multilingual or other) risk groups. Here, it seems, language experts would be well advised to overtly specify the limits of generalizability imposed by the nature of any scientific endeavor. Along this line of thought, the most important part of proper methodological groundwork is the insight of the inevitable constriction of the area of application of any empirical evidence and of the genuinely unstable nature of scientific knowledge in general.

Acknowledgements

References

Atteslander, Peter, Christiane Bender, Jürgen Cromm, Busso Grabow, and Gisela Zipp 1991 *Methoden der empirischen Sozialforschung.* Berlin/New York: de Gruyter.

Beijering, Karin, Charlotte Gooskens, and Wilbert Heeringa 2008 Predicting intelligibility and perceived linguistic distance by means of the Levenshtein algorithm. In *Linguistics in the Netherlands*, Marjo van Koppen, and Bert Botma (eds.), 13–24. Amsterdam: John Benjamins.

Bergman, Manfred Max 2008a Advances in mixed methods research: theories and applications. Los Angeles: Sage.

Bergman, Manfred Max 2008b The Straw-Men of the Qualitativ-Quantitative Divide and their Influence on Mixed Methods Research. *Advances in mixed methods research: theories and applications*: 11–21.

Berthele, Raphael 2011 On abduction in receptive multilingualism. Evidence from cognate guessing tasks. *Applied Linguistics Review* 2: 191–220. Berlin/ New York: de Gruyter.

Berthele, Raphael, and Amelia Lambelet 2009 Approche empirique de l'intercompréhension: répertoires, processus et résultats. *LIDIL*: 151–62.

Blaikie, Norman W. H. 1991 A Critique of the Use of Triangulation in Social-Research [May]. *Quality & Quantity* 25: 115–36.

Bortz, Jürgen, and Nicola Döring 2003 *Forschungsmethoden und Evaluation*. 3. Auflage. Berlin: Springer.

Braunmüller, Kurt, and Ludger Zeevaert 2001 Semikommunikation, rezeptive Mehrsprachigkeit und verwandte Phänomene. Eine bibliographische Bestandsaufnahme. Hamburg: SFB

538. Bryman, Alan 1992 Quantitative and Qualitative Research: Further Reflection on their Integration. Ashgate, Aldershot, England. In *Mixing Methods. Qualitative and Quantitative Research*, Julia Brannen (ed.), 57–78. Aldershot; Brookfield, USA: Avebury.

Denzin, Norman Kent 2005 *The Sage handbook of qualitative research*, 3rd edn. Thousand Oaks Calif. [etc.]: Sage.

Flick, Uwe 2008 *Triangulation: eine Einführung*, 2. Auflage (Qualitative Sozialforschung Bd 12). Wiesbaden: VS, Verlag für Sozialwissenschaften.

Glaser, Barney G., and Anselm L. Strauss 1967 *The discovery of grounded theory: strategies for qualitative research*. 9th pr. (Observations). Chicago: Aldine Publ.

Gullberg, Marianne, Peter Indefrey, and Pieter C. Muysken 2009 Research techniques for the study of code-switching. In *The Cambridge Handbook of Linguistic Code-switching*, Barbara E. Bullock, and Jacqueline A. Toribio (eds.), 21–40. Cambridge: Cambridge University Press.

Hammersley, Martyn 2008 Troubles with triangulation. In *Advances in mixed methods research*, Manfred Max Bergman (ed.), 22–36. London: Sage.

Heeringa, Wilbert, Peter Kleiweg, Charlotte Gooskens, and John Nerbonne 2006 Evaluation of String Distance Algorithms for Dialectology. *Proceedings of the Workshop on Linguistic Distances. Sidney, Austra¬lia Association for Computational Linguistics*: 51–62.

Hufeisen, Britta, and Nicole Marx 2007 EuroComGerm – Die sieben Siebe. Germanische Sprachen lesen lernen. Aachen: Shaker Verlag.

Kaiser, Irmtraud, and Elisabeth Peyer 2011 *Grammatikalische Schwierigkeiten beim Lesen in Deutsch als Fremdsprache. Eine empirische Untersuchung*, vol. 25 (Perspektiven in Deutsch als Fremdsprache). Hohengehren: Schneider.

Lado, Robert 1957 Linguistics across cultures: Applied linguistics for language teachers. Ann Arbor: University of Michigan.

Nunan, David 2008 *Research methods in language learning*. Cambridge: Cambridge University Press.

Odlin, Terence 1989 Language Transfer: cross-linguistic influence in language learning. Cambridge: Cambridge University Press.

Peyer, Elisabeth, Irmtraud Kaiser, and Raphael Berthele 2010 The multilingual reader: advantages in understanding and decoding German sentence structure when reading German as an L3. *International Journal of Multilingualism* 7: 225–239.

Popper, Karl Raimund 2004 *Alles Leben ist Problemlösen: Über Erkenntnis, Geschichte und Politik*, JubiläumsEd. (Serie Piper 4122). München: Piper.

Ringbom, Håkan 1990 Effects of transfer in foreign language learning. In *Current trends in European second language acquisition research*, Hans W. Dechert (ed.), 205–218. Clevedon: Multilingual Matters.

Seliger, Herbert, and Elana Shohamy 1990 *Second language research methods*. Oxford: Oxford University Press.

van Lier, Leo 1988 *The Classroom and the Language Learner*. London: Longman.

Waldmann, Michael R. 2002 Experimente und kausale Theorien. In *Das experimentalpsychologische Praktikum im Labor und WWW*, Dietmar Janetzko, Michael Hildebrandt, and Herbert A. Meyer (eds.), 43–100. Göttingen: Hogrefe.

Weinreich, Uriel 1953 *Languages in contact. Findings and Problems*. The Hague/Paris: Mouton.