The Cambridge Handbook of Linguistic Code-switching

Edited by

Barbara E. Bullock

and

Almeida Jacqueline Toribio

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521875912

First published 2009

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Cambridge handbook of linguistic code-switching / edited by Barbara E. Bullock
and Almeida Jacqueline Toribio.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-521-87591-2
1. Code switching (Linguistics) I. Bullock, Barbara E. II. Toribio, Almeida
Jacqueline, 1963–
P115.3.C36 2008
306.44–dc22

2008026924

ISBN 978-0-521-87591-2 hardback

2

Research techniques for the study of code-switching

Marianne Gullberg
Peter Indefrey
and
Pieter Muysken

2.1 Introduction

The selection of techniques to include in a review of methods for studying code-switching (henceforth CS) inevitably raises the question of what constitutes CS. In addition to the distinctions made between CS and code-mixing (see Muysken 2000), a further distinction can usefully be made between CS and so called language switching.

The term CS can be reserved for studies where the focus is on internally generated switches, i.e. switches produced spontaneously by a multilingual speaker. Studies of CS are mainly concerned with the nature of the constraints governing CS and its processing. The focus is typically on phrases or sentences, on the semantic or structural relationships, and on the linguistic constraints governing switching. The methods used to study CS defined this way are almost exclusively found in the domain of free production.

In contrast, language switching studies generally concern the mechanism of switching itself, including language selection, control, attention, and switching costs, the structure of and access to the bilingual lexicon, and bilingual memory (see Costa 2005; Costa and Santesteban 2006; Dijkstra 2005; La Heij 2005; Meuter 2005). The general domain is the lexicon at the single word level. Typical of language switching paradigms is the use of externally induced switching whereby participants change language on an external cue or respond to an externally generated switch. Frequent techniques are word recognition and lexical decision tasks.

The overarching methodological problem regarding experimental techniques is how to study CS without compromising the phenomenon, i.e. how to induce, manipulate, and replicate natural CS. Arguably, CS as defined above is a production phenomenon whereas the comprehension of CS can only be studied using language switching techniques, i.e. methods

relying on external cues or materials with pre-existing switches to which bilinguals must respond. This raises methodological issues. Although some experimental methods use internally generated switches to study CS, externally generated switches (i.e. language switching techniques) are often used to study both language switching and CS. Language switching techniques can naturally be used to gain insights into CS, but it is important to validate such techniques to ensure that data from language switching methods are actually comparable to CS in the sense of internally generated switching.

The importance of this point can be underscored by considering attempts to study CS using intuition data. Some of the earlier studies on syntactic properties of CS used this technique (Di Sciullo et al. 1986; Mahootian 1993; Woolford 1983). While there may be some degree of correspondence between intuition data and naturalistic data, there is no guarantee that the results coincide. Sobin (1984) systematically studied judgements of adjective/noun orders in Spanish–English CS, where the orders in both languages often differ (*the* ***white*** *house* versus *la casa* ***blanca***). In addition to considerable disagreement between speakers, the overall tendencies reported only partially correspond to what is actually found in naturalistic data. Reliability may vary between different speech communities and different constructions, but is certainly not a given.

It is further crucial to stress that different methods and techniques allow different questions to be answered. Depending on whether the focus of interest is on language switching or on CS, on sociolinguistic aspects, grammatical constraints, phonetic properties, development, on-line processing, bilingual memory, the cost of switching, or the neurocognitive underpinnings of CS, different techniques must be employed. The tension between naturalistic, ecologically valid approaches and more artificial, controlled, experimental techniques should be recognized but also be embraced as a source of complementary information rather than as a (false) dichotomy between "good" and "bad" approaches to the study of CS. A more important consideration in the study of CS is that techniques are validated, and that the behavior they elicit in some measure corresponds to the phenomenon "in the wild" (see Myers-Scotton 2006b). Converging evidence from a range of techniques is perhaps the best way to ensure such validity.

2.2 Naturalistic data and corpus methods: the limits of the observational paradigm

Before the development of good recording techniques made it possible to gather naturalistic data, CS could be studied either through written observations of naturalistic speech (a technique used by Schuchardt 1890; Weinreich 1953), or through the study of texts containing CS. Incidental observations of naturalistic speech have the advantage of at least reflecting

actual language use, but may be inaccurate in various ways. No representative sample can be gathered this way. The study of written texts containing CS can take various forms:

(1.) Studying literary authors who represent the CS of their characters. The best known of these is Tolstoy, who represents Russian–French CS in *War and Peace* (Timm 1978). A number of studies have examined literary texts, but the same reservations apply as for written observations.
(2.) Studying poetic genres where bilingual language use and CS are central to the genre itself. In modernist "high" literature, examples would be some of the *Cantos* of Ezra Pound or portions of *Finnegans Wake* by James Joyce. More amenable to systematic analysis of CS are various types of bilingual poetry, ranging from the Hebrew–Arabic–Spanish *kjarkas* of medieval Spain (Zwartjes 1997) to Yiddish–Hebrew–Russian mixed songs (Weinreich 1950, alas only available in Yiddish), to the *calypso* of Trinidad and the bilingual Quechua–Spanish *wayno* of the Andes (Muysken 2005). The trouble with these bilingual genres is (i) that some switches may be due to requirements of meter or rhyme, and hence do not represent naturalistic speech; (ii) there may be more outrageous mixing than is ever encountered in naturalistic speech, for poetic effect (see Muysken 1995).
(3.) Studying archival material (e.g. trade registers) in which two languages are systematically used (e.g. Middle English and Latin). The drawback of these materials is that they are often quite formulaic, and only represent a limited range of constructions.

To overcome these disadvantages, many researchers, starting at least with Poplack (1980), record naturalistic data in a variety of settings: the public domain, peer group interactions, family gatherings, sociolinguistic interviews, classroom interactions, etc. The work of the last thirty years has produced an impressive corpus of bilingual speech illustrating CS and a host of other language interaction phenomena. Clearly the field owes whatever achievements have been attained to this technique.

Why, then, not simply continue using this method as the main data gathering technique? The naturalistic data method has a number of drawbacks:

(1.) **Costs**. Gathering and transcribing a large corpus of bilingual speech is complicated and costly.
(2.) **Accountability**. For various reasons (competition between researchers, the privacy of the bilingual speakers recorded, incomplete or fragmented transcription, negligence), virtually none of the bilingual corpora on which the CS studies are based are publicly available. It is therefore not possible to study the same materials in order to test the conclusions reached or explore other interpretations.

(3.) **Inherent limitations**. A corpus of naturalistic data has inherent limitations. Some questions are very difficult to answer on the basis of a corpus. These are not only questions that generative syntacticians might ask (e.g. Is it possible to switch in a clause with a parasitic gap?), but fairly standard research questions as well.

We will illustrate the final point with the example of Finnish–English, perhaps surprisingly the most systematically studied language pair with respect to CS. A number of excellent studies exist, listed in Table 2.1.

Table 2.1 Schematic overview of the studies on Finnish–English code-switching

Author	Date	Location
Lehtinen	(1966)	Northern US
Poplack et al.	(1987)	Canada
Lauttamus	(1990)	US
Halmari	(1997)	Southern US
Kovács	(2001)	Australia

We will focus here on determiner phrases with their attendant case marking. The following observations are made by the various researchers, with different interpretations given to the material. In the Finnish examples the following abbreviations are used: PRT = partitive case; IL = illative case; ACC = accusative case; PL = plural; SG = singular; PAST = past; IN = inessive case; AD = adessive case; IMP = imperfect tense; FS = Finnish stem marker.

(1.) In many switched clauses the English noun is marked with the appropriate Finnish case:

(1) Kerran sä olit pannu si-tä mun ***lunchbox***-iin
once you had put it+PRT my lunchbox-IL
"You had once put it in my lunchbox."

(Halmari 1997: 59)

Halmari (1997) interprets this pattern as consistent with a syntactic-theoretical account (the Government Constraint of Di Sciullo et al. 1986): the Finnish verb *pannu* 'put' carries a Finnish language marker, requiring its direct object to be Finnish, as signaled by the case marking on *lunchbox*. In the view of Poplack et al. (1987) the inserted noun *lunchbox-iin* is instead a borrowing, not a code-switch. Finally, Kovács (2001) claims that the phenomenon is a case of "smooth" switching.

(2.) In many cases there is also a deictic element preceding the switched element, both of which are marked for the same case:

(2) Sit se jätti sen sinne ***library***-in
then it left it-ACC there-IL library-IL
"Then she left it in the library."

(Halmari 1997:59)

(3) Molemmat niinku teki ton ***language***-in koulussa
both-PL as/like do-PAST3SG that-ACC language-ACC school-IN
"Both liked the language at school."

(Kovács 2001:152)

Halmari (1997) suggests that the deictic element is a language carrier, requiring that *library* also be marked with a Finnish case, while in the view of Poplack et al. (1987) the phenomenon is a sign of flagging. Kovács (2001) explains it as a characteristic of new information in dialectal spoken Finnish.

(3.) In almost one third of the data, the case marking is missing on the inserted element. Sometimes a case-marked pre-nominal deictic is present:

(4) Siell-o-n iso intiaanimuseo siell-ä ***Prescott***-Ø
there-AD-is big Indian.museum there-AD
"There is a big Indian museum in Prescott."

(Halmari 1997:64)

(5) Se sai semmose-n ***stroke***-Ø
s/he get-IMP3SG like-ACC stroke
"She had like a stroke."

(Kovács 2001:153)

Again, Halmari (1997) explains the phenomenon by saying that the language marking is carried by the deictic element. Poplack et al. (1987) again evoke flagging. Kovács (2001) explains this as fully marked, flagged switching.

(4.) In some cases, there is considerable pausing:

(6) Nekin pelaa sitä ... pel- jotain ***softball***-i-a
They-also play-3SG that-PRT play-some-PRT softball- FEM.SG.PRT
"They also play softball."

(Kovács 2001:191)

All researchers interpret this as an indication of flagged CS.

(5.) Sometimes, however, the flagging element is not present:

(7) Partitive case missing
Teeksää vai ***Irish-Irish coffee***-Ø?
Make- Q-you ... or
"Are you making Irish coffee or what?"

(Halmari 1997:64)

(8) Mä ostan yleensä ***spruce***-Ø taikka ***Douglas pine***-Ø
I buy-1SG usually spruce or Douglas pine
"I usually buy spruce or Douglas Pine."

(Kovács 2001:153)

According to Halmari (1997) this lack of marking is characteristic of attrition among second generation speakers. Poplack et al. (1987) view it as an example of a non-smooth noun switch, whereas Kovács (2001) explains it as L2-oriented switching.

Even in a case like this, where extensive research has been carried out, observational techniques will not help us answer all the questions, and more specific controlled experiments are called for to determine the contributions of the various factors involved.

2.3 Experimental data and methods

Controlled or experimental methods for studying CS beyond the single-word level are still relatively rare, especially if the distinction between language switching and CS is maintained. Semi-experimental techniques range from controlled elicited production tasks to grammaticality or acceptability judgement tasks. Experimental techniques, usually conducted in the laboratory, involve strict designs with a balanced set of experimental and control conditions and test and filler items. Statistical treatment of the results can be found in corpus-based and experimental paradigms alike.

Experimental methods can be divided into *off-line* techniques, where no time constraint is involved and participants can reflect on their responses as long as they like, and *on-line* techniques, where the time course of language processing itself is at stake. These different methods rely on different measures or dependent variables. Off-line methods often draw on metalinguistic judgements or written production. On-line methods typically involve measuring response/reaction times, and accuracy or error scores. Response times are assumed to reflect processing difficulty such that the longer the reaction time, the more difficult something is to process compared to a control item. By inference, the more difficult it is to process an item, the less grammatical, acceptable, or accessible it is assumed to be. Other indicators of difficulty, especially in production, are hesitations, disfluencies, and repetitions. In both types of techniques, features whose influence is examined and manipulated include the nature of the switched element (e.g. a content vs. a function word), the location of a switch, language choice, proficiency and dominance, word frequency, ambiguity, the preceding sentence context, and the nature of the immediately preceding response. Bilingual and CS experimentation calls for very careful design to control for all relevant factors.

In the following, methods will be briefly presented with a short description of the task, its underlying logic, and the measures it yields. For more detailed descriptions, references are given to example studies in Table 2.2. A final remark must be made about the distinction between

Table 2.2 Experimental tasks and example studies

Task	Example studies
Phonetic/phonological level	
Phonetic categorization/ categorical perception	Bürki-Cohen et al. (1989)
Gating	Grosjean (1980, 1988), Li (1996)
Lexical level	
Cued shadowing	Li (1996), Bates and Liu (1996)
Lexical decision	
Visual	Beauvillain and Grainger (1987)
Auditory	Soares and Grosjean (1984)
Naming	
Simple naming	Macnamara (1967a), Meuter and Allport (1999)
Translation	Altarriba and Mathis (1997), Kroll and Stewart (1994)
Stroop	Altarriba and Mathis (1997), Hamers and Lambert (1972)
Picture naming	Costa and Santesteban (2004)
Word association	Taylor (1971)
Sentence level	
Grammaticality/acceptability judgements	
Written	Sobin (1984)
Auditory	Aguirre (1985), Sobin (1984)
Contrastive pairs	Toribio (2001b)
Content judgements	Kolers (1966), Macnamara and Kushnir (1971)
Sentence matching	Dussias (2001)
Reading times	Altarriba et al. (1996), Dussias (2003)
Auditory moving window	Ferreira et al. (1996)
Reading aloud	Grosjean and Miller (1994), Kolers (1966), Toribio (2001a)
Free speech in switch mode	Azuma (1996), Grosjean and Miller (1994), Toribio (2001a)
Sentence repetition, elicited imitation	Azuma and Meier (1997), Clyne (1972), Meijer and Fox Tree (2003)
Sentence completion	Dussias (2002)
Sentence recall (priming)	Meijer and Fox Tree (2003)
Confederate scripting	Kootstra et al. (in prep.)
fMRI, PET	Hernández et al. (2001)
ERP	Moreno et al. (2002)

comprehension and production tasks. Experimental tasks are generally speaking more numerous for the study of comprehension than production (see Bock 1996). While intuitively convenient, the distinction between comprehension and production is not always easy to uphold since much language activity involves both modalities. Wherever possible, a note will be made regarding whether a task focuses mainly on comprehension or production.

2.3.1 The phonetic–phonological level

To probe the role of phonetic–phonological knowledge in bilinguals' comprehension and production of CS, language switching paradigms have

mainly been used, i.e. tasks measuring responses to externally generated switches (but see Bullock, this volume).

2.3.1.1 Phonetic categorization and categorical perception

Phonetic categorization tasks are used to examine whether bilinguals categorize CS words as belonging to language A or B and what factors influence such categorization. Listeners hear a range of speech sounds forming a continuum between two unambiguous endpoints with ambiguous sounds in between. In CS research an endpoint is a sound or a word clearly belonging to language A or B. The ambiguous sounds are often synthesized or digitally manipulated versions of the endpoint sounds. Listeners are required to identify all sounds as being either language A or B. The task measurements are proportions of responses categorized as A or B, reaction times, and accuracy scores on a companion discrimination task of the endpoints. This task has been used to investigate whether bilinguals mainly use phonetic information or also preceding context to decide what language a CS word is (Bürki-Cohen et al. 1989).

2.3.1.2 Gating

Gating is used to examine how much and which information is necessary for listeners to identify words in speech (Grosjean 1980). For CS, the role of phonetic and phonotactic properties of words, the preceding context, word frequency, etc., can be examined. Typically, participants hear sentences with a spoken target word presented in segments of increasing duration (increments of between 30 and 50 milliseconds) until the entire word has been presented. For each presentation, participants must identify or guess the word (and/or the language) and give a confidence rating of their guess. The task yields measures of the number of segments, or gates, required for participants to reach the isolation point (i.e. the amount of exposure needed for identification), confidence ratings, and the word guesses, which can be characterized for language.

2.3.2 The lexical level

Methods that target individual words are almost exclusively language switching techniques and can be argued to tap general bilingual processing rather than CS.

2.3.2.1 Cued shadowing

Cued shadowing (or single-word shadowing, auditory naming, word repetition) (Bates and Liu 1996) is also used to study the recognition of code-switched items in bilingual speech. The task allows the influence of phonological, structural, and contextual information on the recognition of CS words to be examined. Listeners are presented with a spoken phrase or sentence, which contains a target word that they must repeat as quickly

and accurately as possible once they detect it. The task yields reaction times and accuracy scores. In monolingual contexts, the target word can be recorded in another voice. Li (1996) adapted the task for CS research by presenting the target word in the other of the bilingual's languages (Chinese or English). Participants were told beforehand where in the sentence structure the target word would occur. A variation on this task is to indicate that the target word will start with a particular sound (a phoneme-triggered task, suggested by Heredia and Stewart 2002).

2.3.2.2 Lexical decision

Lexical decision tasks are used extensively in studies on language switching to examine issues of lexical access, the structure of the bilingual mental lexicon, and control issues. Lexical decision tasks examine what determines how long it takes listeners to decide whether a written or a spoken string is a word or not. Participants hear or read words, presented in isolation or in sequence. For a given target word, they must decide whether the string is a real word or not by pressing a button for Yes or No. The response measures are reaction times and accuracy or error scores. The logic of the task is that, if bilinguals consider words in both languages for every decision, it should take them longer than monolinguals to decide whether a string is a word or not. If, in contrast, only one language is active, then response times should be the same as for monolinguals.

The basic task can be modified in many ways. Written versions manipulate orthographic properties of words. For instance, issues of lexical access are studied by measuring the influence of words that are orthographically similar but mean different things in two languages (interlexical or interlingual homographs, "false friends," Beauvillain and Grainger 1987). Orthographic systems with both overlap and differences, as in Greek and Latin scripts, are also manipulated (see Dijkstra 2005 for an overview of studies of visual word recognition). In auditory versions participants are asked to listen for a string starting with a particular sound in a sentence and then to indicate whether that string is a word or not (phoneme-triggered lexical decision, e.g. Soares and Grosjean 1984).

In cross-modal versions, e.g. cross-modal priming, participants listen to speech (the "prime," e.g. *dog*) while watching a fixation mark on a screen. A written test word then appears on the screen, which can be semantically (e.g. *cat*) or phonologically (e.g. *doll*) related or unrelated (e.g. *apple*) to the prime. Participants must then make a lexical decision or name the word. Reaction times are measured to determine whether priming occurs, i.e. whether the prime word led to facilitation (shorter reaction times) or to interference (longer reaction times).

2.3.2.3 Naming tasks

Naming tasks are often used in the study of language switching to examine the structure of the lexicon, the strengths and nature of associations

between words of the same versus different languages, and the cost of switching language. Naming tasks are used to study both comprehension and production. The simplest version requires participants to name as many words as possible in a given time window using a particular language or a particular mode, for instance, to name only in one language, to switch language for every word, or to give a translation equivalent for every word offered (Macnamara 1967b). The dependent variable is generally the number of words produced and response times. Monolingual naming is compared to switched naming. The logic is that the strength of association between words, levels of activation, and facilitation/inhibition of lexical access will be reflected in the number of words produced.

Other versions constrain the task by asking participants to name particular stimuli and by providing an external cue for switching. For example, participants may be asked to name digits with the language to be used indicated by a geometrical shape (Macnamara et al. 1968), background color (Meuter and Allport 1999), or a particular digit system (Campbell 2005). A translation version requires participants to provide the translation equivalent of a keyword, varying the directionality of the translation from language A to B or vice versa (e.g. Chan et al. 1983; Kroll and Stewart 1994). In these tasks, naming latencies and accuracy or error scores are measured. Participants may also be required to read monolingual or mixed word lists out loud in which the switching frequency is manipulated (Dalrymple-Alford 1985). Naming latencies in such studies are measured to examine whether switching language takes longer than not switching. A further variation is a translation recognition task in which participants are presented with two words, one in each language. They are required to decide whether the second word is an accurate translation of the first by pressing a button for Yes or No (e.g. Altarriba and Mathis 1997). Again, reaction times and accuracy/error scores are measured.

A particular type of naming task is the bilingual Stroop task (MacLeod 1991). Participants must name a color patch (e.g. in red) on which are printed either incongruently colored words (*blue*), congruently colored words (*red*), or neutral words (*car*). The task is to name the color of the patch, and to ignore the words. The difference in response times between naming patches with incongruent words and neutral words is known as the interference effect. In bilingual versions color words appear in both languages, and participants are asked to respond in language A on some trials and in B on other trials. The question is whether it is easier to ignore the printed word when it is in another language than the response language, e.g. to ignore the printed word *"red"* printed on a blue patch, when you must respond "azul" in Spanish. Although interference is typically greater within a language, there is also a robust interference effect between languages. Interference is also likelier from a dominant into a weaker language than the opposite (see Altarriba and Mathis 1997;

MacLeod 1991). The Stroop task can also be auditory. Listeners may be asked to decide whether the pitch of a speaker's voice is high or low as the speaker pronounces the words "high" or "low" in either a high or low pitch and in alternating languages (see Hamers and Lambert 1972).

Another type of naming is picture naming. The simplest kind presents pictures to be named on different colored backgrounds to cue the use of a particular language (e.g. Costa and Santesteban 2004). The picture names are manipulated to examine the effect of cognates, for instance. The measures are reaction times and error/accuracy scores. In picture-word interference tasks a picture is presented on a computer screen and participants are asked to name the picture (e.g. a dog) as quickly and accurately as possible. They are also asked to ignore an auditory or visual interfering stimulus that can be semantically (e.g. *cat*) or phonologically (e.g. *doll*) related to the picture, or not related at all (e.g. *apple*). In bilingual versions the distractors can come from both languages, and the naming can take place in either or both of the bilinguals' languages (see La Heij 2005 for an overview). Again, the rationale is that if naming in language A is affected by the names of distractors in language B, this is an indication that access to the bilingual lexicon is not language-specific.

2.3.2.4 Word associations

Word association tasks in spoken and written form are used to study the relationship and associations between words in the bilingual lexicon and bilingual memory. Participants are asked to provide word associations to keywords in their two languages. They can be requested to respond monolingually, in either language, or to switch at particular rates, such as once every two words or at every other word (e.g. Ervin-Tripp 1964; Taylor 1971). The dependent variable is the number of words provided. The task can also be used to probe whether bilinguals will make semantically or conceptually different associations in either of their languages.

2.3.3 The sentence level

Only at the sentence level are techniques found that draw on internally generated switches to study CS, as opposed to language switching paradigms where the switch is externally imposed. Both written and auditory stimuli are used to elicit written and spoken output.

2.3.3.1 Grammaticality or acceptability judgement tasks

Traditionally, grammaticality and acceptability judgement tasks are written off-line tests probing participants' grammatical knowledge. In bilingual studies, participants must respond by indicating whether a sentence with a particular type of switch is grammatical or not, or indicate its degree of acceptability on Likert scales (e.g. Bhatia and Ritchie 1996; Sobin 1984). To reduce the risk of prescriptive attitudes towards switching,

auditory versions have been developed in which participants instead hear the sentences to be judged (e.g. Aguirre 1985; Sobin 1984), and instructions can be phrased in familiarity terms, for instance, "Does this sentence sound like something you might have heard?" (Aguirre 1985). In addition to the judgement, participants are occasionally asked to correct or improve unacceptable sentences (e.g. Lederberg and Morales 1985). Another variation presents sentences in contrastive pairs, drawing attention to switching sites (Toribio 2001b). This procedure can fruitfully be combined with scalar responses or be turned into a preference response rather than an absolute one to avoid some of the artifacts and validity problems connected with the task (e.g. Sorace 1996). The technique can be used to examine the acceptability of particular switch locations (e.g. Toribio 2001b).

2.3.3.2 Content judgements

A more indirect measure of acceptability are content judgements such as comprehension questions or sentence verification tasks calling for true–false judgements. The assumption here is that if a switch of a particular type (e.g. manipulation of switching site or element switched) incurs a processing difficulty, then response times should be longer and error rates higher. Such tasks are typically language switching tasks since the materials are pre-switched and artificially created. Comprehension questions are often off-line tasks where participants respond after silently reading or listening to monolingual and code-switched passages of text or speech (e.g. Kolers 1966). True–false judgements can be on-line tasks in which participants read or hear a monolingual or code-switched passage. They are then asked to judge whether the sentence is true or false by pressing a button for Yes or No. The task yields reaction times as well as accuracy/error scores (Macnamara and Kushnir 1971; Rakowsky 1989; Wakefield et al. 1975).

2.3.3.3 Sentence matching

Sentence matching is an on-line task that also taps acceptability indirectly without asking explicit metalinguistic questions (Freedman and Forster 1985; Gass 2001). Participants see two written sentences on a computer screen, one presented slightly after the other. They must indicate whether the two sentences are identical or not by pushing a button for Yes or No. The time needed for this same–different decision is measured, as are accuracy/error scores. The pacing of presentation time can be fixed or self-paced, in which case participants themselves bring up the second sentence by a button push when they have read the first one, the so-called response-contingent sentence matching technique. The first technique provides only one response time, the second provides one response time for reading the first sentence, and another for the matching decision. Generally, the first reading time is given most weight, as it is assumed that reading time reflects complexity of processing. In general, grammatical

sentences are responded to faster than ungrammatical sentences. A promising modification for CS research would be an auditory version of the task (see Roberts and Verhagen forthcoming). In CS research, Dussias (2001) has used the written task to investigate the effects of switch locations in Spanish–English, comparing switches before function words like determiners (***la maestra compró*** the books "the teacher bought ...") to switches between function and content elements (***la maestra compró los*** books "the teacher bought the ..."). Switches at dispreferred locations or in dispreferred directions (from English to Spanish or from Spanish to English) are assumed to affect response times.

2.3.3.4 Silent reading

Reading is used to tap processing in language switching. Reading times are assumed to reflect processing difficulty such that the longer the reading time, the more difficult the processing. The unit for which reading time is measured ranges from whole text passages to sentences, phrases or individual words. Eye-tracking techniques allow one to measure the time the eye fixates on an individual word during reading. The number of times an element is fixated on can also be counted, again assuming that more fixations indicate greater processing difficulty (see Dussias 2003; Altarriba et al. 1996). The technique can be used to examine reading times of switched words in mixed sentences.

2.3.3.5 Auditory moving window

In this task participants listen to sentences one or two words at a time and must press a button to receive successive segments (see Ferreira et al. 1996). The time needed to process each segment is recorded. The technique has been employed to study the effect of context, word frequency, and phonetic realization in the processing of switched sentences (Heredia et al. 2002).

2.3.3.6 Reading aloud

A simple sentence-level language switching technique is to ask bilinguals to read texts aloud that are either monolingual or that contain switches (e.g. Chan et al. 1983; Grosjean and Miller 1994; Kolers 1966; Toribio 2001a). The measure is always reading (plus speaking) speed and the aim is to investigate the influence, if any, of switching itself. Switching can be "blocked" with switching occurring at alternating sentences, or before a particular word class (e.g. each noun). Switching can also be random but conform to the word order of one of the languages, or it can be entirely random. The contrasts between acceptable versus unacceptable switches have also been examined. The task has also been used to assess whether bilinguals change to language-appropriate phonetics at the critical word or not by measuring voice onset times (see Bullock et al. 2006).

2.3.3.7 Free speech in "code-switch mode"

A straightforward CS task is to ask bilinguals to speak freely over a given topic, either monolingually in language A or B, or to deliberately code-switch. Although speakers are required to code-switch, they are free to decide when and how to do so. Switches are therefore internally generated and constitute true CS. The assumption is that the procedure will yield "natural" switches compatible with the bilinguals' grammars. The measure is the number or proportion of switches produced per some speech unit (per clause, per 100 words, etc.), as well as more qualitative measures like types of switches, switch locations, phonetic-phonological properties of switches, etc. The degree of fluency can also be measured. Speech can be entirely free (Blot et al. 2003), consist of story retellings in speech or in writing (Grosjean and Miller 1994; Toribio 2001a), or be spoken summaries of texts that are either monolingual or switched (Kolers 1966). This method generates ecologically valid data, but does not allow much control over factors that constrain CS. A language switching version of the task asks bilinguals to speak freely over a given topic and then to switch language when they hear a tone randomly generated at irregular intervals. With this method Azuma (1996) examined the planning units of switched speech. He noted the type of word uttered at the time of the tone (noun, preposition, etc.), measured the time it took the bilingual to stop speaking and switch language, and noted the type of element where language was switched (word, phrasal boundary, etc.).

2.3.3.8 Sentence repetition

In sentence repetition or elicited imitation participants hear a sentence and must then repeat it back as accurately as possible. The rationale of the technique is that when listeners hear a sentence that exceeds the capacity of their short-term memory, they will pass it through their own grammar before repeating it. If a particular grammatical element in the input sentence is not part of the individual's grammar, that element will be changed during the repetition (Vinther 2002). The dependent measures are therefore number or proportion of accurately repeated sentences, alternatively, number or proportion of changes, the time elapsed between the offset of the prompt and the onset of the repetition, and qualitative aspects of types and locations of changes. Because the nature of switches and their locations is predefined, it is typically a language switching task. However, the repetitions produced may be considered to be CS. The delay between hearing the sentence and the required repetition can vary from immediate repetition (Azuma and Meier 1997), to a ten-second delay with intervening music (Clyne 1972), to delays involving distractor tasks (Meijer and Fox Tree 2003). The longer the delay, the more likely the response is to reflect the individual's own grammar.

2.3.3.9 Sentence completion

In sentence completion tasks, participants are presented with sentences with missing elements. Participants must read and complete the sentence by filling in the blank as quickly and accurately as possible. The time between the last read element and the naming of the blank or of a corresponding picture can be measured as a response time and accuracy/error scores can be recorded. Properties of the filled-in element can also be recorded, like language chosen, word order, grammatical category, etc. The task relies on internally generated switches, i.e. on CS proper, since no force is applied to make the bilingual switch. If a switch occurs, it is self-generated. The task can be used to examine, for instance, preferred switch locations such as between functional and content elements.

2.3.3.10 Sentence recall (priming)

Sentence recall tasks combine comprehension and production aspects. Participants are asked to read and memorize a sentence. They then read a prime sentence with a different syntactic construction, and/or in a different language. After reading the prime, participants are asked a distractor question, for example, whether a certain word was part of the prime sentence. Finally, they are asked to recall the original sentence aloud. The key question is whether the structure of the prime will influence recall of the original sentence. The task has been used to examine syntactic priming effects, i.e. whether listeners are more likely to repeat syntactic constructions they have just heard than to use another construction (Hartsuiker et al. 2004; Loebell and Bock 2003). For example, Meijer and Fox Tree (2003) used the opposition between double-object and prepositional object constructions in English and Spanish to examine whether a construction in one language would prime the corresponding construction in the other. These studies are not CS studies *per se*, but could be modified for the study of CS.

2.3.3.11 Confederate scripting

A related technique is confederate scripting (Branigan et al. 2000; Hartsuiker et al. 2004) where two participants take turns describing a picture. The confederate participant is instructed (scripted) to use a particular lexical or syntactic construction. The extent to which the real participant then uses the same construction in his/her own description is measured. Kootstra et al. (in preparation) use the task to elicit Dutch-English CS. Participants are given a lead-in sentence fragment like "on this picture" and are asked to describe a picture, switching language somewhere in the sentence. The confederate produces code-switched sentences that conform either to Dutch or English word order given the lead-in fragment. The confederate's contribution constitutes a language switch but the real participant's output is CS. The dependent measure is

the number of times the real participants use a primed word order, which word order is used, and where in the structure the switch occurs.

2.3.4 Neurocognitive methods

Recently neurocognitive methods have been used to investigate the neurocognitive underpinnings of bilingual processing (see papers in Gullberg and Indefrey 2006). Both hemodynamic (PET, fMRI) and electrophysiological (ERP, MEG) techniques are promising research tools for the study of language switching and CS. Normally, participants need to perform no additional experimental task other than reading or listening for comprehension. With some limitations due to possible movement artifacts, the techniques can also be used to investigate language production. A number of studies have examined the neural correlates of language switching using explicit cues to switch language (see Rodriguez-Fornells et al. 2006 for an overview). Electrophysiological methods are very sensitive to the time course of neural events in relation to a particular word. The technique therefore provides precise measurements of the effects of a code-switch on processing and is also suited to examine neural events preceding a code-switch in production.

The ERP technique provides a means for distinguishing semantic from syntactic difficulties, and for assessing task control and attention. The monolingual literature has already established signatures for semantic integration difficulties. When a participant has difficulties integrating the semantics of a word with the preceding context, this is typically reflected in an increased negativity peaking approximately 400 ms after the onset of the unexpected word in comparison to a condition where there is no integration difficulty. This is the so-called N400 effect (Kutas and Hillyard 1984). Similarly, difficulties integrating syntactic information are reflected in an increased positivity peaking approximately 600 ms after the onset of the critical element, the so-called P600 effect (Osterhout and Holcomb 1992; Hagoort et al. 1993). These signatures can be exploited to study the effects of CS on bilingual lexical and sentence processing. A recent study investigated the electrophysiological responses of English–Spanish bilinguals to sentences and idioms that ended either in the expected English word, in an English synonym of the expected word (a lexical switch) or in a Spanish translation equivalent (a code-switch) (Moreno et al. 2002). See Kutas et al. (this volume) for an expanded discussion of ERP research in CS.

2.4 A multi-task approach to studying CS production

The above review of experimental techniques highlights the challenges involved in choosing tasks that ensure ecological validity of experimental

data, that take the sociolinguistic and contextual sensitivity of CS into account, and that acknowledge the tension between tasks operating with externally and internally generated switches, here labeled language switching versus CS (see Grosjean 1998). One possibility is to use multiple tasks to probe the same phenomenon and to generate converging evidence. This is not a new idea. For instance, in an early study, Ervin-Tripp (1964) used word association tests, sentence completion tasks, semantic differentials, and so-called problem stories in the assessment of Japanese–English bilingual speech. Recently, Toribio (2001a) has combined the reading aloud of manipulated texts with spoken and written story retellings in a forced output mode to assess code-switching competence. Similarly, Dussias (2002) has examined the processing of a particular structure in both comprehension and production using eye tracking in reading and a sentence completion task.

In a research project targeting the study of code-switched sentence production, researchers can fruitfully combine multiple methods and collect data from the same participants performing a variety of tasks. The procedure allows for several types of baseline and within-subject comparisons. Baseline data are necessary to establish what constitutes natural CS in a particular population and language pair (see Muysken 2000 for a typology of different switching types in different bilingual communities). The experimental tasks and the data they yield can further be validated against this baseline. This approach allows researchers to examine whether the patterns of sentence-level CS observed "in the wild" are replicable experimentally, which is a worthy research goal in itself.

An example of this approach is a project that targets Papiamentu–Dutch bilinguals living in the Netherlands (Gullberg et al. in preparation). Papiamentu is a creole language of mixed Spanish–Portuguese origin spoken in the Dutch Antilles, where it co-exists with Dutch, which has a colonial past as an official language (e.g. Gordon 2005; Kouwenberg and Muysken 1995). The Papiamentu–Dutch bilinguals in the Netherlands code-switch as a normal part of everyday interactions (see Muysken et al. 1996; Vedder et al. 1996).

In a study of CS, the effect of language mode must be controlled for (Grosjean 1998). Bilingual experimenters therefore recruit and instruct participants in code-switched mode, except when the tasks call for monolingual performance, in which case different, monolingual experimenters are called upon. The tasks move gradually from interactive, multiparty settings where CS is more likely to occur, to individual settings and tasks; from spontaneous and naturalistic to controlled and experimental tasks; and from tasks where no constraints are applied to make participants code-switch, to more constrained tasks. A list of sample tasks is presented in Table 2.3.

The conversations and the Director–Matcher task are entirely unconstrained. The Director–Matcher task is a referential communication task

Table 2.3 Tasks and output modes

Task	Language output mode
Conversations	Free (incl. CS)
Director–Matcher (CS)	Free (incl. CS)
Sentence completion task with picture naming	Free but stimulated CS
Shadowing	Constrained CS
Auditory acceptability judgement task	Constrained CS
Director–Matcher (Dutch)	Constrained mono Dutch
Standardized Dutch proficiency test	Constrained mono Dutch

(Yule 1997) in which two participants have to solve a problem together. One of them has the information necessary to solve the task and must convey it so that the other participants can "match" the information and thereby solve the task. Although the task can be designed to encourage speakers to use particular constructions, they are not coerced to use a particular language. The sentence completion task is also free but stimulates switching. The shadowing and acceptability judgement tasks are fully constrained.

The output from the various tasks can be compared qualitatively and quantitatively to examine the validity of the more constrained tasks. For example, lexical noun phrases (NPs) with a modifying color adjective of the type *the white flag* could be targeted. Within-NP switches are of particular interest because they represent potential conflict sites given the word order properties of the two languages: Dutch adjectives are prenominal, *de witte flag*, and Papiamentu color adjectives are typically postnominal, *e bandera blanco*. The Director-Matcher data – e.g. elicited items such as "arrange *the green bottle* first" – may be compared to the conversation data to determine whether the switches from the former task are "natural." In this way, the data from the Director-Matcher task constitute a baseline for complex NPs and possible within-NP switches against which to validate the experimental items in the sentence completion task.

The sentence completion task is designed to examine whether the language of the first sentence element, the subject-NP, or the finite verb will influence the naming of the direct object. In this task speakers are not forced but implicitly encouraged to switch. The experimental items are transitive sentence frames consisting of a subject-NP and a finite transitive verb. These are presented as text and the direct object is a colored picture to be named as quickly and as accurately as possible. The lead-in sentences can be either monolingual Dutch or Papiamentu (9–10), or have the subject NP in one language and the finite verb in another (11–12). The participants read the sentences and name the picture in whatever language they choose. The task is designed to examine whether the language of the first sentence element, the subject-NP, or the finite verb will influence the naming of the direct object.

(9) *Papiamentu*
e homber den e bar ta kibra [picture of green bottle]

(10) *Dutch*
de man in de bar breekt [picture of green bottle]

(11) *Papiamentu–Dutch CS*
e homber den e bar ***breekt*** [picture of green bottle]

(12) *Papiamentu–Dutch CS*
de man in de bar ***ta kibra*** [picture of green bottle]
"The man in the bar breaks [picture of green bottle]."

If constructions similar to those from the Director–Matcher task are attested, the sentence completion task may be said to yield qualitatively ecologically valid output. In addition, these tasks can be complemented with a battery of more controlled tasks, such as shadowing (Marslen-Wilson 1973) and auditory acceptability judgement. Output from these tasks allows for a comparison of bilinguals' explicit metalinguistic judgements with their on-line processing of a particular structure.

2.5 Conclusions

Research techniques employed in the study of CS have clearly progressed in their methods of data collection, mirroring the progress made with other phenomena of psycholinguistic inquiry. The earliest studies focused on individual and incidental observations. Later studies involved increased attention to the importance of relative frequencies in the observed data. Most recently, CS studies involve the careful variation of experimental conditions in controlled, laboratory settings. Despite the recent trend toward more experimental techniques, it should be clear from this chapter that, even though naturalistic data have their limits, experiments can never fully replicate or replace observations of naturalistic CS. There are benefits to be gained from integrated studies that seek to validate experimental methods and data against naturally occurring CS.