An investigation of control processes in bilinguals, using visual word recognition tasks

Thesis submitted to the University of London
for the degree of Doctor of Philosophy

Roswitha E. von Studnitz

Department of Psychology
University College London
December 2001

ACKNOWLEDGEMENTS

Over the past years many people have contributed to the work locked inside these two blue covers (with some cousins in journals). The contributions and the people have been of many different kinds. From the far-flung family endless encouragement - and questions of what progress has been made since the last week.... But also simple daily telephonic questions how *I* am (rather than my work). What a welcome respite from my carrousel of thoughts. Dearest Maika,dearest Paika - how can I ever thank you enough? Without you, this would not have been possible. Luka, Raban, Alix, Freya: not only have you been a moral support, you helped me to get the ratings for my experiments, asking around among friends and (student) colleagues. What a help that was! Your efforts meant not only obtaining the ratings that allowed me to put together my experimental stimuli, but it spared me from having to 'waste' precious German-speakers here in London, whom I had to save for the experiments themselves. From Brussels/Paris/Königswinter/Brussels - and finally London! - a remedy to keep one sane: an alternative to the seemingly endless round of preparation, running, evaluation of experiments and writing up: a life! Thank you, Andrew (T.O.), for your unfailing good-will. To you all, my dears, enormous, heartfelt thanks! Our various countries' telephone companies will be wondering why their profits are down now... Debbie - from willing participant at very German early-morning hours to dear friend - I thank you. Colleagues of mine were astonished at (and envious of) how early I could begin running experiments in the day... Shaheen, what a friend and 'mommy' you have been - many thanks for believing in me when I didn't - and for your cakes! Fahimeh, I am sorry we only began to be in contact at a later stage - to you also many thanks for your kind thoughts and wishes. Clive, the world of computers can be such a puzzle - for help with this, and *much* else in the final stage: many, many thanks! Annette, ich habe die Öhrchen steifgehalten, wenngleich sie manchmal etwas schlaff hingen. Dino, thank you for the prolonged computer use in room 423! To several hundreds of German/English bilinguals who participated: none of this would be here without you. Though you were 'remunerated at the standard departmental rate' (initially an enormous £3 per hour, then increased to a staggering £5 per hour!), I realise and appreciate the time and energy each of you gave me essentially out of good-will. Simon, how could I have ever managed to get my experiments and evaluations going without you? Indeed, how would I have climbed the final hurdle without your kind help? I hope you know how much your help meant (and means) and how grateful I am. Dr. Green, or as even I with a very German view of university hierarchies learned to say: David, for many, *many* times of discussion, and getting-back-on-track, thank you very much. To Professor Stephen Monsell and Dr. Matthew Saxton, many thanks for making the viva experience one that I think back to with interest and composure, rather than with a shudder... (*That's enough thank-yous. - Ed.*)

PUBLICATIONS BASED ON THIS THESIS

von Studnitz, R.E. & Green, D.W. (1997). Lexical decision and language switching. *International Journal of Bilingualism*, **1**, 3-24. (Experiments 1a/1b)

von Studnitz, R.E. & Green, D.W. (2002). Interlingual homograph interference in German/English bilinguals: Its modulation and locus of control. *Bilingualism: Language and Cognition*, **5**, 1-23. (Experiments 4a/4b)

von Studnitz, R.E. & Green, D.W. (2002). The cost of switching languages in a semantic categorization task. *Bilingualism: Language and Cognition*, **5**, part 3 (in press). (Experiment 3)

PRESENTATIONS BASED ON THIS THESIS

R.E. von Studnitz & D.W. Green, *The cost of switching languages: The effects of varying the lexical decision task.* International Symposium on Bilingualism, Newcastle upon Tyne, 9th - 12th April 1997.

R.E. von Studnitz & D.W. Green, *Interlingual homograph effects in lexical decision.* International Symposium on Bilingualism, Newcastle upon Tyne, 14th - 17th April 1999.

R.E. von Studnitz & D.W. Green, *Dynamic control of interlingual homograph interference in (German/English) bilinguals.* ESCoP XI, Ghent, 1st - 4th September 1999.

ABSTRACT

How is competition between bilinguals' languages resolved? This thesis contrasts two main accounts - an *internal* (Bilingual Inhibitory Control [BIA] model, Dijkstra & van Heuven, 1998) and an *external* account (Inhibitory Control [IC] model, Green, 1998). The internal account proposes varying the activation of languages within the lexico-semantic system; the external account proposes task schemata acting differentially to the signals emerging from this system. On-line visual word recognition tasks (with German/English bilinguals) measured RTs and errors.

Experiments 1a/1b: switching languages incurred a greater cost when language was task relevant than when it was not ('language-specific' vs. 'language-general' lexical decision task), consistent with both external and internal accounts. **Experiment 2** sought to distinguish between the two accounts by refining Experiment 1. In a language-specific lexical decision task language-switch costs were reduced when this task followed the language-general task, but only for 'pure-HF' stimulus lists, not for 'mixed' (HF / LF) lists. Such patterns are consistent with an external, but not an internal account, and provide insight into lists' effects on decision criteria. **Experiment 3:** animacy decisions (for German and English words) produced a language-switch cost only in the first half of the experiment, and only for response-repetition trials, not for response-switch trials. This is consistent with an external, but not an internal account, and shows the influence of language signals on decisions even when task-irrelevant. **Experiments 4a/4b:** English-specific lexical decisions incurred a cost on interlingual homographs (vs. matched controls). The degree of interference was greater if pure German words (to be rejected) were included than when none were included. This enhanced interference diminished with time (Experiment 4a). Interference was less when indviduals were informed about the presence of IHs than when not informed (Experiment 4b). IH interference and 'carry-over' effects dissociated, consistent with an external but not an internal account. **Overall**, the data are consistent with a non-selective access account. They further indicate that language activation levels can be raised exogenously, but not reduced endogenously. A mechanism reacting differentially to signals from the lexico-semantic system can account for the observed patterns. This mechanism appears to be influenced by task demands, experience, item constraints and available information. The external account is parsimonious, proposing the same mechanism for resolving competition between languages, as for control of action.

In the text the 3rd person singular 'he' is used from time to time, which should be taken to mean both 'he' and 'she'. The studies undertaken here were given ethical permission by the Joint UCL/UCLH Committees on the Ethics of Human Research: Committee Alpha.

TABLE OF CONTENTS

I Introduction and review: Language processing in bilinguals

II Experiments 1a & 1b: Language-switch costs in language-specific and language-general lexical decision tasks

III Experiment 2: The effects of task order and list composition on language-switch costs in language-specific and language-general lexical decision tasks

IV Experiment 3: Using bilingual animacy decisions to contrast an internal and an external account of language-switch costs

V Experiments 4a & 4b: The locus of controlling interlingual homograph interference

VI General Discussion

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS USED IN THIS THESIS

ADT	animacy decision task (Experiment 3)
ANOVA	analysis of variance
B&M	Battig &Montague (1969) word typicality norms (Experiment 3)
BIA	Bilingual Ineractive Activation model (e.g., Dijkstra & van Heuven, 1998)
C	matched control word (Experiments 4a/4b)
CELEX	Centre for Lexical Information (word frequency norms for English, German & Dutch; Baayen, Piepenbrock & van Rijn, 1993)
CS	contention scheduling
HF	high frequency word
HF1	the first high frequency word following a specific item (Experiments 4a/4b)
HF2	the second high frequency word following a specific item (Experiments 4a/4b)
IBM	International Business Machines
IC	Inhibitory Control model (e.g., Green, 1998)
IH	(non-cognate) interlingual homograph (Experiments 4a/4b)
K&F	Kučera & Francis (1967), word frequency norms
L1	first (in some cases dominant) language
L2	second (in some cases subordinate) language
LD	lexical decision (Experiments 1a/1b & 2)
LDT	lexical decision task (Experiments 1a/1b & 2)
LF	low frequency word
LLEX	Lexical tests (English, German and other language vocabulary tests, Meara, 1994)
ms	milliseconds
MSE	mean square error
NW	nonword
PC	personal computer
PFC	prefrontal cortex
RT	reaction time
SAS	supervisory attentional system
SCT	semantic categorisation task (Experiment 3)
SD	standard deviation
SOA	stimulus onset asynchrony

I

Introduction and review: Language processing in bilinguals[1]

1.1 Introduction

Two German/English bilinguals spent a day conversing exclusively in English, as was their habit. Late in the evening, when considering where to leave the car for the night, one said to the other: "Let's park the car over there, it's nice and /bɹaɪt/ there, I mean, it is *wide* there". The word /bɹaɪt/ was pronounced as an English word, i.e., as 'bright'. When this word is pronounced in German (which differs only slightly from the English pronunciation of the word 'bright'), it is the German word 'breit' (/bʁaɪt/), which means 'wide'.

This bilingual, who normally spoke English fluently, and had spoken only English during the entire day, suddenly experienced interference from German. The bilingual did not simply 'borrow' a German word, to insert into an English sentence ("it's nice and breit there"), which would have conveyed the intended meaning. Instead, this was clearly an example of an interlingual interference, which disrupted the intended message. Dornič (1978) and Clyne (1980) suggest that when tired, distracted, or unwell, bilinguals may experience interference between their two languages, even if they do not under normal circumstances. If it is possible for bilingual language processing to be 'impaired' from time to time, how is such impairment/interference avoided normally? Or in other words, how is the bilingual language processing system normally regulated?

Interesting evidence about disruption to bilingual language processing also comes from cases of individuals with brain damage[2]. From the domain of cognitive neuropsychology, there are reports that previously competent bilinguals may lose control over their languages. For example, patient AD (Paradis, Goldblum & Abidi, 1982), previously a fluent French/Arabic bilingual, presented with a distinctive pattern of language impairment following a head injury. She was able to speak French but not Arabic on one day, with a

---

[1] The term 'bilingual' when used in the following, is taken to mean a person with a high level of competence in both languages; this includes cases where both languages were learned at the same time and cases where they were learned consecutively. The participants taking part in the experiments presented here almost all fall into the latter category.

[2] "The inner workings of a complex system are often revealed by the way in which the system breaks down." (Dell, 1986, p. 284).

reversal on the following day ('alternate antagonism'). Compounding the first problem, when spontaneous production in French was possible, she could not translate into French (although she could translate into Arabic, which she could not produce spontaneously); this deficit too was reversed on the following day ('paradoxical translation'). (See also Albert & Obler, 1978; Fabbro & Paradis, 1995.) Patient AD is interesting, as her case highlights the need for an explanation involving the *functioning* of the language processing system. An explanation referring only to the *representation* of words is inadequate, as one cannot argue that impairment of AD's ability to speak French derived from her 'French language system' having been impaired. After all, a language that she could not speak on one day, was available on the following day, after which it would disappear again, and so on. Furthermore, a language that she could not speak spontaneously, she could produce when translating.

What underlies interlingual interference in unimpaired bilinguals, and how is it possible for this interference to be variable? How are alternate antagonism and paradoxical translation possible? To answer such questions, one needs to understand how the bilingual language system functions.

Early studies of bilingualism tended to ask how a bilingual's two languages are represented, leaving aside the question how the representational system functions (e.g., Weinreich, 1953; Ervin & Osgood, 1954). The views expressed by these early models generally fell into one of two camps: Languages are represented in two independent systems, to which access is selective; the other view was that languages are represented in interdependent systems, to which access is nonselective (i.e., occurs in parallel). A problem with these models (described in section 1.5) is that they were able to accommodate either 'independent' performance[3], or 'interdependent' performance[4], but not both. This is clearly a shortcoming, particularly given that behaviour consistent with the languages being independent, and behaviour consistent with the languages being interdependent, may exist within the same individual (see introductory example).

Furthermore, if it is the case that a bilingual's two languages are connected, then should there not be constant interference between them? How does a proficient bilingual

---

[3] e.g., Scarborough, Gerard & Cortese (1984): bilinguals at times fail to recognise an unexpected 'other language' word.

[4] e.g., Preston & Lambert (1969): In a bilingual Stroop task, bilinguals were affected by both intra- and interlingual word/colour discrepancies (see end of section 1.5.2.1).

manage to avoid interference between his two languages? The examples given above suggest that there is some means of control, which averts interference, but which may fail at times. The question then is, how is this control achieved? That is the topic of this thesis.

This thesis examines the notion of an 'external' control mechanism (e.g., Green, 1986; 1998a, b) which regulates interlingual interference (in language perception) from outside the bilingual language processing system. This external model is contrasted with an 'internal' control mechanism, that regulates interlingual interference by modulations within the language processing system (e.g., Grainger & Dijkstra, 1992; Dijkstra & van Heuven, 1998). Both models are described in detail in section 1.6.

At present the notion of a control mechanism is comparatively new. It needs to be tested and better understood. Achieving this would provide the basis for gaining a better understanding of disorders of AD's type (and thus be able to develop suitable therapeutic intervention - this is further discussed in Chapter 6). Given that this approach to investigating bilingual language processing is relatively recent, several fundamental issues need to be addressed. For example, to study the processing of whole sentences, it is necessary to have an understanding of what is involved in the processing of single words. Consequently, this thesis focuses on single word processing. Specifically, the experiments presented here required competent adult bilinguals (German / English) to process visually presented single words, under various task instructions.

1.1.1 Summary: why 'control' is an important issue

Striking clinical cases, such as the patient AD mentioned above (Paradis, Goldblum & Abidi, 1982; but see also Albert & Obler, 1978; Fabbro & Paradis, 1995), point to a control mechanism which may break down, affecting the ability to use a language system which may otherwise be intact. Secondly, in unimpaired, competent bilinguals, interlingual interference may occur from time to time.

Thirdly, bilinguals use both their languages intentionally; in a controlled, not haphazard, manner. To translate text, the bilingual must avoid simply reading the text. In speaking, bilinguals may speak only one language, or switch from one to another at will. There must be some mechanism which controls this apparent ability to select output in language A when input was in language B (when translating). For spontaneous speech, there must be some means of avoiding the accidental insertion of a word from the other language[5], which might not be understood by the listener (one language must be selected, and maintained as such). For intentional switching between languages[6], there must be some means of changing between output in language A and output in language B (there must be a shift from having one language 'selected' to having the other language 'selected').

There is also the less obvious issue of what effect language switching has when it concerns input. For example, how is mixed-language input processed, i.e., what happens in the language processor when the language of input changes? To understand language input, the incoming stimulus must be matched to an existing mental representation of that word. As discussed in detail in the second half of this chapter, there is evidence that there are separate word form representations for each language, to which access is parallel (rather than serial) and which are interdependent (rather than independent). This type of set-up could explain interference between languages. However, how does such an arrangement accommodate the finding that it takes longer to process mixed language input compared to pure-language input (e.g., Kolers, 1966; Macnamara, 1967; Macnamara, Krauthammer & Bolgar, 1968; Macnamara & Kushnir, 1971; Thomas & Allport, 1995)? If the two language systems are interconnected, and sensory information is fed into them simultaneously, then why should switching languages incur a temporal cost? The assumption pursued here is that there is some means of control, which allows the system to behave in what appears as a language-selective or language-non-selective manner, depending on the circumstance. The question is, how and where this control is achieved. This is the question addressed by this thesis.

An explanation for this co-existence of language-selective and language-non-selective behaviour cannot be found simply with reference to representation, since the two phenomena suggest opposite explanations. The first suggests separate representations, the latter integrated, or at least interacting, representations. Looking for an answer to this seeming contradiction in the *functioning* of the bilingual language system, is more successful, as shall be shown in this thesis.

1.2 Overview

To provide a background for the questions addressed here, Chapter 1 first furnishes an outline

---

[5] "I could go out now, oder I could go later" ('oder' is German for 'or')

[6] "Shall we take the tube? Ich denke das wäre am schnellsten" (I think that would be quickest)

of the unilingual lexico-semantic system[7] (Votaw, 1992). This approach is adopted on the basis that it is simpler to describe for unilinguals, and that the fundamental principles of language representation are similar in unilinguals and bilinguals (e.g., Paradis, 1977; Beardsmore, 1980; Grosjean, 1982; Kirsner, Lalor & Hird, 1993). Then, a review of the literature pertaining to bilingual language processing is provided, focusing on language perception, which is the area relevant to the research presented here. Using this background, the question about function, how bilinguals control language processing, can be addressed. Starting from early models (e.g., Macnamara & Kushnir's 1971 'input-switch' model), the development of selected models is sketched, up to two key contenders: the 'internal control' and the 'external control' models. An important aspect of these two is that they, for the first time, directly address how the system may be controlled. The internal account is exemplified by Dijkstra & van Heuven's (1998) Bilingual Interactive Activation (BIA) model. The external account is exemplified by Green's (1998a, b) Inhibitory Control (IC) model. The ensuing experiments (Chapters 2 to 5) were designed to test these two models. Where relevant, some early 'representational' models are re-assessed. Chapter 6 summarises and discusses the key findings.

1.3 The unilingual lexico-semantic system

This thesis asks how bilinguals control their lexico-semantic system, not how words are represented in bilinguals. Nonetheless, selected aspects of language representation and processing shall be outlined here for various reasons. 1) The experiments presented in this thesis require word recognition, so it seems necessary to address the underlying representations and processes at least to some degree. 2) In the remainder of the thesis references will be made to 'lexical representations' and 'semantic representations'[8]; this section serves to give these terms some meaning. 3) The BIA model was developed from an earlier unilingual model (McClelland & Rumelhart's [1981] Interactive Activation model), so it is important to describe those underpinnings here.

---

[7] Lexico-semantic system: the mental system in which words and the parameters defining them (e.g., word form, pronunciation, syntactic properties, meaning) are represented.

[8] Note, in this thesis the term 'semantic' does not include 'conceptual' representations. Conceptual representations (accumulated connotations associated with a word, not simply its meaning), are regarded as a separate level of representation (for a discussion on the importance of making this distinction, see Pavlenko, 1999; see also Paradis, 1979, 1997; Grosjean, 1998). The representations which the experiments of this thesis focus on are lexical and semantic, but not conceptual. Hence this latter level shall not be defined further here.

The view that shall be outlined here is that the lexico-semantic system includes some means of analysing letter features (e.g., horizontal lines, angles, curves), some representation of letters, of words, of word properties (e.g., syntactic properties), of meaning, and of the sounds of words (which allows print to be turned into sound).

1.3.1 The Interactive Activation model - the basis of the Bilingual Interactive Activation model

An influential account of word recognition is the Interactive Activation (IA) model of McClelland & Rumelhart (1981). This model is relevant insofar as it was the basis for developing the Bilingual Interactive Activation model (Grainger & Dijkstra, 1992; Dijkstra & van Heuven, 1998), which is one of the two models tested in this thesis. Note that the IA model accounts for the recognition of four-letter words, which clearly is only a small proportion of what humans are capable of.

In the IA model there are three levels of representation: for features, for letters, and for words. Each level stores information as a 'localist' representation. There are connections between the representations (or 'units') at a given level, as well as between units at different levels. These links allow other units to be activated or inhibited. The feature level detects letter features and activates all letter representations containing that feature (e.g., detecting a discontinuous left hand curve activates C and G), all other letter representations must be inhibited (e.g., O, Q, D). Letter units activate compatible word representations, while inhibiting all incompatible ones. Activation is restricted to certain representations, in that feature and letter representations are coded for position. Thus perceiving a word-initial C activates CAT but not ICE[9]. 'Word neighbours' (words identical to the target word bar one letter[10]) receive some activation as well, since they have aspects that are compatible with the input. Following Coltheart et al. (1977), it is assumed that access to relevant lexical candidates occurs in parallel, rather than in a serially ordered search process (e.g., Forster, 1976, 1979). Since the target word's lexical representation matches the presented letter string best, it receives most activation. Receiving most activation, this representation can rise above

---

[9] But note that there are word-similarity findings which suggest that positional coding is not very strong. Chambers (1979) found that, in a lexical decision task, TRIAN is more difficult to reject than TRUAN; yet, as regards letter positions, both are equally dissimilar to TRAIN, posing difficulties for a positional coding account.

[10] Coltheart, Davelaar, Jonasson & Besner, 1977; e.g., neighbours of CAT are CAR, CUT, BAT, etc.

a threshold level of activation[11], surpassing other potential candidates, and can thus be selected (cf. Grainger, 1992; Johnson, 1992; for a comparable theory for acoustic word perception, see Luce & Pisoni, 1998).

There are some established factors about language processing that the IA model does not cover. Those relevant to the issues addressed in this thesis are outlined next.

The IA model assumes that word recognition is determined by visual information; yet there is evidence that sounds of words can influence word recognition as well. For example, in a lexical decision task[12], nonwords that *sound* like words (e.g., brane) are more difficult to reject than 'true' nonwords (e.g., brene) - see Morton (1996). Secondly, reading a word (e.g. 'frog') is facilitated not only if a previously read letter string is a related word ('toad'), but also if it is a nonword that sounds like a related word ('tode') - Lukatela & Turvey (1994). Such results suggest that, upon letter identification, both orthographic and phonological representations are activated, both of which have access to semantic representations.

McClelland & Rumelhart's (1981) IA model does not account for meaning storage or retrieval (the same applies for the early versions of the BIA model). Meaning is language's raison d'être, so this is an aspect that needs to be considered. Meaning representations of words are addressed in section 1.3.2 (below).

At the beginning of this section it was noted that apart from the word's form and meaning, there must also be some way of representing its syntactic properties. McClelland & Rumelhart's (1981) IA model does not clearly address this aspect. Levelt (e.g., Levelt, 1989; Bock & Levelt, 1994) allocates this role to 'lemmas'. He proposes that (in a network model), there is a lemma level of representation that stores the syntactic properties of represented words (e.g., CAT = noun). For an individual who speaks more than one language it is important to know which language a word belongs to and what its syntactic properties are in that language. Language affiliation may also be coded at the lemma level (see Green, 1998a). For example: 'CAT: it is an English word, it is a noun, its determiner is neutral (the)'; 'KATZE: it is a German word, it is a noun, its determiner is feminine (die)'.

---

[11] Note, the point at which a word's (localist) representation is recognised is characterised in different ways by different models. Morton's logogen model (e.g., 1969) is based on the assumption of different thresholds for each representation (the higher the word's frequency, the lower its threshold). McClelland & Rumelhart's IA model (1981) is based on the assumption of a common threshold of activation, but different 'resting levels' or 'baseline levels' of activation (the more frequent the word, the higher the baseline activation). Gordon's resonance model (1983, 1985) is based on the assumption of different rates at which words can gain activation (the higher the frequency, the faster the gain).

[12] "is the letter string is a word or a nonword?"

Another problem with the IA model is that it does not speak clearly to the question of new word acquisition (how does one develop a new lexical 'unit'?). This is something distributed models address more easily (e.g., Seidenberg & McClelland's [1989] 'distributed developmental model of word recognition and naming'). In distributed models, recognising a word does not involve the identification of a single word unit, or node, as localist models presume. Instead, word recognition involves reinstating a pattern of activation distributed over several nodes and the connections between them. In Seidenberg & McClelland's (1989) distributed model, there is no position coding for letters as in the localist IA model (McClelland & Rumelhart, 1981). Instead, letters are coded in triples, such as _CA, CAT, AT_ for the word CAT (_ denotes an empty position). These triples (comparable to Wickelgraphs, see Wickelgren, 1969) are encoded as an activation pattern over a network of units. A word is recognised once the network of units associated with that word has reached a stable state of activation.

Note, this thesis does not address the manner in which words are represented in the lexico-semantic system, it simply takes as given that there is some form of a representation which is activated following (visual) presentation of a stimulus. Therefore the debate over localist vs. distributed representations shall not be pursued here. Let it suffice to say that the distinction between 'localist' and 'distributed' is not as clear as it might seem: Localist models generally assume some degree of distributed representation as well, in that while words are represented in a 'localist' manner at the lexical level, word representations are 'distributed' at the lower levels (i.e., in terms of letters and features) - cf. Page (2000).

1.3.2 'Dissociations' and cognitive representations

Much pertinent evidence about the unilingual language-processing system has come from cognitive neuropsychology, where it was shown that impairment could be restricted to certain abilities. Special cases of such selective impairment[13], as well as double dissociations[14], played a significant role in formulating models of the language processing system. Examples relevant in particular to Experiment 3 pertain to animacy/inanimacy of things around us (e.g.,

---

[13] or 'single dissociation', meaning that patient X is impaired in task A but not in task B.

[14] patient X is impaired in task A but not in task B, whereas patient Y is impaired in task B but not task A; e.g., different types of aphasia: intact word order with concurrent omission of inflections (Tissot, Mounin & Lhermitte, 1973), impaired word order concurrent with accurate use of inflections (Saffran, Schwartz & Marin, 1980).

dog vs. house), and the differentiation between lexical and semantic representations (i.e., the word-form DOG vs. the meaning {domestic animal, furry, four-legged, barks}).

For instance, Warrington & Shallice (1984) found impairments in naming, verbal description, object recognition, miming and picture-word matching of animate but not of inanimate entities, suggesting some difference between the way in which the two are represented semantically (see Johnson-Laird, 1987). Regarding lexical and semantic representations of words, Funnel & Allport (1987) and Schwartz, Marin & Saffran (1979) found cases of people able to distinguish between real words and nonwords, without being able to understand the real words' meanings, suggesting that lexical and semantic information have separate cognitive representations. There is converging evidence for this view from experiments with healthy individuals. For instance, Besner, Smith & MacLeod (1990) showed participants two letter strings simultaneously, asking them to determine whether they had a letter in common ('letter matching task'). Decisions were faster for strings that were real words than for strings that were not, suggesting lexical involvement (rather than processing the task at a letter level of representation). Critically, letter matching was slower when the two letter strings were semantically related words (e.g., bread - butter), than when they were unrelated (e.g., bingo - butter). In short, semantic relatedness and lexicality had opposite effects on task performance, suggesting separate representations of lexical and semantic information. (See also Barron & Henderson [1977]; Henik, Friedrich & Kellogg [1983]; Smith [1979]; Smith, Theodor & Franklin [1983], who used other paradigms to address this notion; see Smith [1997] for a discussion.) This notion of separate lexical and semantic representations is of importance for the experimental work presented later on (specifically, Experiment 3), in that it provides an opportunity to devise a task that distinguishes between the Bilingual Interactive Activation model and the Inhibitory Control model, which are tested here (further details are in section 1.6.1).

The suggestion that lexical and semantic information are represented separately does not entail that processing at, for example, the lexical level is uninfluenced by semantic factors. Indeed, there is evidence that lexical decisions are influenced by semantic properties (for a discussion see Balota, 1994), which suggests that lexical processing is not isolated from semantic processing, and that accessing semantic information may not be strictly consecutive to accessing lexical information (cf., for example, Morton, 1996). In accord with the notion that semantic properties of a word can influence processing in a 'lexical' task (a task that theoretically would require information only from the lexical level), Whittlesea & Cantwell (1987) showed that nonwords provided with a meaning showed a word superiority effect. Using a different approach, Kellas, Ferraro & Simpson (1988) demonstrated faster reaction times to homographs than to nonhomographs (e.g., organ vs. bugle) in a lexical decision task (stimuli were controlled for potential confounding factors such as familiarity). Gernsbacher (1984) and Chumbley & Balota (1984) provided converging evidence from yet another angle: Lexical decision times were found to be sensitive to concreteness, a semantic variable.

1.4 Experimental tasks and levels of processing

Two tasks frequently encountered in the literature of language experimentation are lexical decision tasks and semantic categorisation tasks (an instance of which is the animacy decision task). These are the two types of task employed in the course of this thesis, and thus they shall be explained briefly.

In a lexical decision task ('LDT' - developed by Landauer & Freedman, 1968; Rubinstein, Garfield & Millikan, 1970), participants need to differentiate between real words (e.g., CAT) and nonwords (e.g., NOF). Finding a lexical representation in the lexicon would theoretically suffice to provide an answer, no further processing is necessary. That is, a signal emerging from the lexico-semantic system indicating 'a lexical representation has been found', would suffice. Apart from identifying a specific representation, it may also be possible to judge that the item in question is a real word rather than a nonword on the basis that there is much general activation in the lexicon (cf., for example, Grainger & Jacobs, 1996). This is further discussed later on in this chapter.

Normal language processing is concerned with conveying meaning. Therefore, apart from activating representations within the lexicon, the corresponding representation at the semantic level would presumably, quite naturally, be activated as well, even though the task does not require it. Activation of a semantic representation would give rise to a signal that also emerges from the lexico-semantic system ('a semantic representation has been found'). Phonological word representations could play a role as well. According to Ellis & Young's (1988) model for the recognition and production of spoken and written words (based on evidence predominantly from cognitive neuropsychology), written words activate not just 'visual' representations, but also phonological representations. Identifying an existing phonological representation would also provide evidence that the letter string is a word.

In short, for a lexical decision task, the ability to determine whether a letter string is a word or a nonword may come from various signals, including signals from the lexicon and the semantic system, that are the most relevant for the tasks presented here.

In an animacy decision task (ADT) participants are asked to differentiate between words pertaining to animate entities and those referring to inanimate ones (e.g., LION and CAR, respectively). For this task, finding a corresponding representation in the lexicon is insufficient grounds for an accurate response. It is necessary to identify the meaning of the letter string, and on this basis decide whether the word in question refers to an animate or inanimate entity (how 'animacy' may be coded is discussed in Chapter 4). From the semantic system the relevant information emerges - this is the crucial signal in order to be able to give a response. Signals from the lexicon will presumably become available in this task as well, indicating that there is a corresponding lexical representation. However, these will not be of use for distinguishing between animate and inanimate (nor would any signals about phonological representations).

To summarise, a task aimed at a specific level of representation does not just 'tap into' that level, but may reflect processing at other levels as well. Secondly, task instructions determine which signals are of use for responding and which are not (e.g., lexical signals are useful given instructions to distinguish between words and nonwords, but not given instructions to distinguish between animate and inanimate entities).

Signals from the various systems need to be organised in some way, in order for a participant to provide an answer in an experimental task. For example, it is not sufficient to recognise a letter string as being a real word. The participant needs, somehow, to convey the knowledge that the item is a real word, to a response mechanism. Only then will he be able to give an overt "yes" or "no" response. For the current purposes, this is envisaged as a sort of 'linking mechanism', which (based on task demands) links appropriate signals emerging from the lexico-semantic system to the units underlying the appropriate response. For example, in a lexical decision task, all signals indicating that the letter string is a word, are linked to the units underlying a 'yes' response, while all signals indicating that it is not a word, are linked to the units underlying a 'no' response. When either one of the response units has achieved dominance over the other, a response can be given. This linking mechanism is described in more detail later on, in conjunction with the bilingual data. Such a mechanism, is, in fact, the locus of control proposed by the external account (e.g., Green, 1998a, b), which is tested here.

Given this background, it is an appropriate juncture to introduce the bilingual language processing system. The first part of this chapter addressed the question why it is important to consider the question how bilinguals control their languages (bilinguals showing behaviour consistent both with 'independent' and 'interdependent' language systems; cases like patient AD [Paradis, Goldblum & Abidi, 1982]). As will become clear in the following review of the bilingual language processing literature, the question how bilinguals achieve such control is a relatively recent one. Following an initial impetus by Green (1986), some researchers have attempted to model how such control might be effected. Two key models shall be described: The Bilingual Interactive Activation model ([internal account] e.g., Grainger & Dijkstra, 1992; Dijkstra & van Heuven, 1998) and the Inhibitory Control model ([external account] e.g., Green, 1998a, b). These two models are important since the control mechanism, which they address, is clearly necessary, yet was not accounted for in any of the early bilingual studies. The following section describes these early investigations and shows how, by focusing solely on the question of representation, they were unable to accommodate behaviour consistent with both 'independent' and 'interdependent' language representations, or to account for the range of abilities bilinguals have (such as translation, or code-switching).

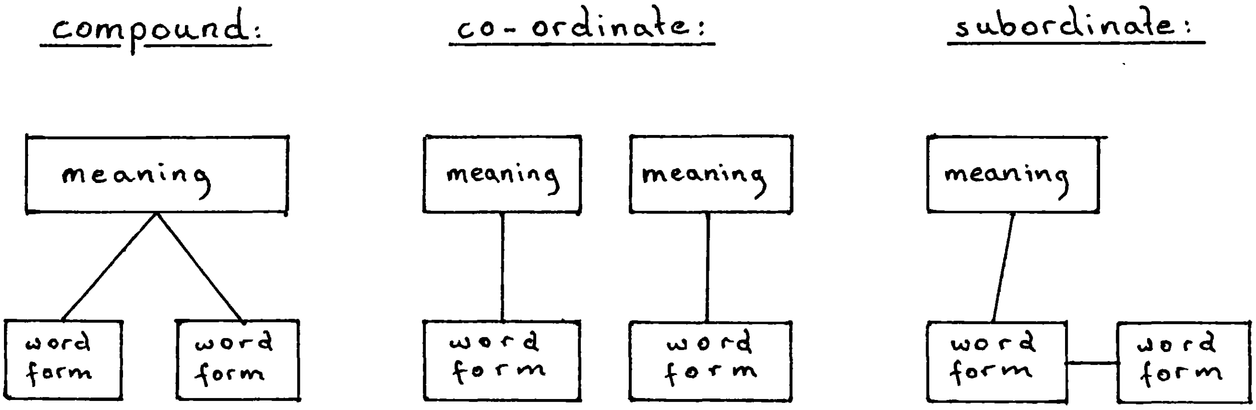
1.5 The bilingual lexico-semantic system.

1.5.1 Early models of the bilingual lexicon: how are the languages represented?

The early models (Weinreich, 1953; Ervin & Osgood, 1954) focused on how the languages are represented, and suggested three possible representational arrangements for the two languages in bilinguals: 1) the Co-ordinate Type, where two separate word-form lexica are each linked with two separate meaning systems; 2) the Compound Type, where two separate word-form lexica link into one unified meaning system; 3) the Subordinate Type, where a weaker language's word-form lexicon is linked to one common meaning store via the stronger language's word-form representations (see Figure 1.1).

Figure 1.1

Co-ordinate, compound, and subordinate language representations in bilinguals (see Weinreich, 1953).



A number of questions arise for these models. How does a compound bilingual manage to select between the two possible lexical (or 'word-form') encodings available for a particular meaning? (For example, a German/English bilingual can express the meaning {cat} either by the word-form CAT or the word-form KATZE ['cat' in German].) Secondly, how does a co-ordinate bilingual manage to translate? With two separate systems, it is not clear how the necessary links between the languages are made, and, concerning the question of control, how the individual is able to select language A output when the input was language B. (For example, having read the word 'cat' and retrieved the meaning {cat}, the bilingual knows what meaning he is meant to convey, but how does he manage to express that meaning as 'Katze'?) In short, Ervin & Osgood's (1954) account leaves open the question of how control is achieved.

The compound/co-ordinate distinction was upheld for a while, on the basis that there appeared to be evidence of culture-specific differences in the mental images triggered by translation equivalents (Lambert, Havelka & Crosby, 1958). For example, the terms 'BROT' and 'PAIN' ('bread' in German and French, respectively) produce different mental images for a German than for a Frenchman. When further research failed to find significant differences between associations produced by the two types of bilinguals (e.g., Dillon, McCormack, Petrusic, Cook & Lafleur, 1973), the individual's language history ceased to be considered a primary determinant of mental language organisation[15]. There was a change in emphasis, whereby the focus of attention turned to issues of language processing.

Experimental methodologies were developed which allowed bilingual language processing to be investigated more scientifically, focusing on statistical validity and controllability of the dependent variables. For instance, the time taken for comprehending passages of text, modified by the number of language switches within the text (e.g., Kolers, 1966); the time taken for true/false decisions for sentences (that are either sensible or nonsensical), again modified by the numbers of language switches within the text (e.g., Macnamara & Kushnir, 1971); synonym judgement task with mixed- and same-language word pairs (McCormack, 1977).

Research building on the pioneering work of Weinreich (1953) and Ervin & Osgood (1954) continued to address the question of representation, still without considering how the system might be controlled. The focus had turned to the question whether words of the two languages access separate cognitive representations (one for each language), or whether they access common, abstract, representations (e.g., Kolers, 1963; McCormack, 1974, 1977). For some time this question remained unresolved, not least because no clear theoretical distinction was made between lexical and semantic representations of words. Some studies focused on the former, others on the latter (Durgunoğlu & Roediger, 1987, and Smith, 1991 evaluate and discuss this problem in detail). In fact, as early as 1979, Dornič noted that it was problematic not to make this distinction. Today, there is generally a consensus that lexical representations are coded for language (i.e., there are separate lexical representations for each language, possibly represented in separate lexica, or possibly represented in an integrated lexicon; see van Heuven, 2000; Kroll & Dijkstra, 2001), while semantic representations appear to be language-neutral (and hence presumably represented in a single, common system); see, for example, Durgunoğlu & Roediger (1987); Kroll & de Groot (1997). This difference has led to the conflicting experimental results and interpretations.

1.5.1.1 Separate lexica, one for each language?

Different languages have different word forms to encode the same meaning (e.g., CAT vs. KATZE ['cat' in German]). To account for this orthographic difference, one presumably must

---

[15] The importance of this factor for language processing was re-introduced later, see Grosjean, 1998.

assume separate lexical representations for CAT and KATZE, just as one would for CAT and DOG (see Grainger & Dijkstra, 1992). How are these words represented in the bilingual? Are there two lexica, one for each language, or are words of both languages represented in one common lexicon? Additionally, each lexical representation must be coded for language in some way, as bilinguals are able to determine whether a word belongs to language A or to language B. This coding could be envisaged, for example, as a language tag associated with each lemma (e.g., Green, 1998a), or as a language node to which all words of a given language are connected (e.g., Dijkstra & van Heuven, 1998). The implication is not two neuroanatomically distinct systems, but of two neurofunctional systems. As shall be discussed next, there does not appear to be clear evidence for the assumption of separate lexica.

The assumption of separate lexica (and a shared semantic store) is made by various researchers (e.g., Scarborough, Gerard & Cortese, 1984; Kroll & Sholl, 1992; Kroll, 1993; Smith, 1997). It is frequently based on the experimental finding of within-language repetition priming, but no between-language repetition-priming (also known as translation-priming). For example, Kirsner, Brown, Abrol, Chadha & Sharma's (1980) Hindi/English bilinguals performed a lexical decision task, divided into two blocks. The language of decision was either the same in both blocks, or different in the two blocks (e.g., block 1: Hindi words and Hindi-like nonwords; block 2: Hindi words and Hindi-like nonwords [within-language condition], or, block 2: English words and English-like nonwords [between-language condition]). Some words were repeated between blocks. A German/English example for a within-language condition is CAT ... CAT; for a between-language condition an example is CAT ... KATZE. The repeated word was responded to faster only in the within-language, not the between-language condition. Kirsner, Smith, Lockhart, King & Jain (1984, Experiment 1) repeated the experiment with English/French bilinguals, which meant using two languages sharing an alphabet. Here too, within-language, but not between-language, effects were found. (See also Scarborough, Gerard & Cortese, 1984.)

However, these results cannot be taken as evidence for separate lexica. The interval between the prime and target was several minutes long. If one regards translation pairs (CAT / KATZE) as the bilingual equivalent of synonyms in unilinguals, then the lack of a translation priming effect could be seen as the equivalent to the lack of long-term synonym priming in unilinguals (e.g., reaction times to WOMAN are unaffected by having seen FEMALE several trials earlier) - see Thomas, 1997. Secondly, significant between-language repetition-priming was found when the prime / target interval was reduced to some hundreds of milliseconds, rather than minutes (de Groot & Nas, 1991). The short-term between-language repetition-priming was found even when primes were masked to hinder integration strategies. This led de Groot & Nas (1991) to conclude that translation equivalents are linked to one another at the lexical level ("lexical links"). Activation, they suggested, could spread from one lexical representation to the other directly, allowing a translation equivalent to receive activation even if a prime were not processed sufficiently to retrieve meaning[16]. De Groot & Nas (1991) suggested, following Forster & Davies (1984), that at long prime/target lags, the spreading activation between lexical representations (which would lead to the translation-priming effect) would have decayed, hence giving the impression of there being no between-language repetition-priming.

Gerard & Scarborough (1989) also employed a long-term between-language repetition design, showing English/Spanish bilinguals words and nonwords blocked for language in a lexical decision task. Their study differed from those by Kirsner and others (above), in that they included non-cognate interlingual homographs, that is, words sharing spelling but not meaning across languages (e.g., RED, meaning 'net' in Spanish). Reaction times to these words were apparently unaffected by the other-language word frequency. That is, in the Spanish block reaction times to RED were slow (in line with it being low frequency in Spanish), while they where fast when the same word appeared in an English block (in line with it being high frequency in English). Such results were considered as meaning that there are separate lexical representations for the words of each language, and furthermore, that there are two lexica, one for each language, which do not interact. Note that this is contrary to the masked repetition-priming evidence provided by de Groot & Nas (1991). Furthermore, Gerard & Scarborough's (1989) finding can be reinterpreted as being consistent with separate but interacting ("interdependent") lexica. This is discussed later on.

In sum, there does not appear to be clear evidence that a bilingual has two separate lexica. However, given that lexical representations are distinct for each language, and that they must be coded for language, it makes sense to envisage functionally separate, but interacting lexica. From now on, when reference is made to separate lexica, this is meant simply in a functional sense, not in an anatomical sense, and allows for the two systems to

---

[16] Where a meaning representation received activation, that representations could, in turn, pass activation to the other-language lexical representation top-down. This semantic route is, in fact, the route suggested by Grainger & Frenck-Mestre, 1998, and is compatible with the Revised Hierarchical Model of Kroll & Stewart, 1994.

interact. If there are two lexica, the next question is how they are related to one another.

1.5.2 Are L1 / L2 lexical representations independent or interdependent, and is access to them selective or nonselective?

This section describes evidence that has been taken as indicative of how words of different languages are represented. Different lines of investigation have given rise to conflicting viewpoints. Some researchers argued for two independent systems, to which access is selective; others for interacting (interdependent) systems to which access is nonselective. The following inspects the data taken to support these conflicting views.

1.5.2.1 Language-switching experiments

This section outlines evidence that switching languages takes time. This 'language-switch cost'[17] was taken as evidence that a bilingual's languages are stored separately, and that access to these stores is selective.

Kolers (1966) asked French/English bilinguals to read different types of paragraphs (of ca. 110 words): written either in one language only, or with languages alternating from sentence to sentence, or with languages mixed randomly within sentences. In a silent reading condition, there was no significant difference between the time taken to read the 'pure' and the 'mixed' passages. In contrast, in a loud reading condition, mixed text was read significantly more slowly than pure text. That is, at least when output was required, there appeared to be a cost for switching languages.

Macnamara & Kushnir (1971) adapted Kolers' (1966) method of silent reading comprehension in order to test their proposal of the switch cost's origin. They presumed that there should be a cost not just for switching languages at output, as Kolers (1966) had suggested, but also for switching languages at input. The idea of the '*input-switch*'[18] was that language input was channelled into one or other of the two lexica (which were considered to be independent of each other). The input-switch was assumed to be set to channel the input

---

[17] For the studies in this section, 'switch-cost' meant the overall time increase when processing mixed- rather than pure-language text. For the experiments of this thesis, which investigate on-line processing rather than overall completion times, 'switch cost' refers to the difference between reactions on language-switch trials as opposed to language-non-switch trials (this difference may be an increase in RT or in errors).

[18] The notion of a 'language-switch' makes its first appearance in the work of Penfield & Roberts (1959), based on Hebb (1958), who suggested that a neurophysiological 'switch' exists which mediates between independent neural sets. See also Macnamara (1967) for early use of the notion of a 'switch'.

into the appropriate lexicon once some language information had been perceived. This process was thought to be automatic, as it is impossible to avoid understanding language input (in contrast to an output-switch that determines the language used at output, which clearly is voluntary). In other words, external stimuli re-set the input-switch. For example, identifying letter combinations orthographically legal in only one language (e.g., -PF- is legal in German but not in English) would set the input-switch towards the corresponding lexicon. If there is no language cue (e.g., words orthographically legal in both languages, such as the German word WIRT), then, given no contextual cues (such as the text being purely German), the input-switch presumably stays in the last-used position, which would mean that the wrong lexicon is searched first on a language-switch trial. In such a case, once it has become clear that there was no match, the input-switch would need to be re-set, and the item searched for in the other lexicon. Such a procedure could explain the momentary failure to comprehend an unexpected 'guest word', a known phenomenon among bilinguals (e.g., Grosjean, 2001).

Macnamàra & Kushnir (1971) suggested that they were able to provide experimental evidence for this input-switch conjecture. Imbalanced[19] English/French and French/English bilinguals silently read mixed-language paragraphs (taken, in part, from Kolers, 1966). Silent, rather than loud, reading was employed so as to isolate the input-switch (Experiment 1). The mean 'language-switching' time, was smaller than that found by Kolers (1966) in his loud reading condition, but it was nonetheless significant. Macnamara & Kushnir (1971) took their result as testimony of having found evidence for an input-switch mechanism (input and output switches were thought to function sequentially, each giving rise to its own cost). In a second experiment, participants were instructed to judge whether or not short sentences, containing one, two or three language switches, were true or false[20]. The greater the number of switches, the more delayed the reaction time (RT).

A problem with this study (and to some extent also with Kolers's [1966] study), is that language switches were distributed fairly randomly, rather than adhering to the phrasal structure of sentences, which would have been a more accurate reflection of natural code-switching in bilingual language (e.g., Clyne, 1980; Poplack, 1980; Sridhar & Sridhar, 1980).

Neufeld (1976) assessed the question whether this language-switch cost was indeed due to switching languages (as Macnamara & Kushnir, 1971, assumed), or whether it was an

---

[19] stronger in their L1, weaker in their L2

[20] Examples: Beaucoup de chiens ont *legs*. Des cheveaux *must* manger. Les oiseaux *have* deux *wings*.

artefact of the syntactic position of the language-switches. To this end, balanced bilinguals were presented with both 'naturally switched' and 'unnaturally switched' sentences. There was an auditory and a visual condition. Macnamara & Kushnir's (1971) findings were replicated in an auditory presentation condition, whereby RTs increased with increasing numbers of language switches. However, and most importantly, when language-switches occurred only at phrase boundaries, the 'switch-cost' disappeared. It seems, then, that if a change of language occurs in an anomalous position, this disrupts the predictive processes involved in normal, skilled reading (e.g., Just & Carpenter, 1987; see also discussion in Pinker, 1994). Such an explanation would suggest that a time-cost arises not because a switch is being re-set, but because the change of language affects the processes within the lexico-semantic system.

In order to avoid the problem posed by syntax, which was inherent in the studies by Kolers (1966) and Macnamara & Kushnir (1971), Dalrymple-Alford (1985) presented participants with word lists, to be read aloud. This way, he argued, it would be possible to ascertain whether the switch-cost found by, amongst others, Macnamara & Kushnir (1971), was really a cost for having switched language, or, whether the pure language / mixed language difference had been an artefact of syntactic anomaly (present only in the mixed, but not the pure, paragraphs) as Neufeld's (1976) data suggested. In Dalrymple-Alford's (1985) first experiment, the switch-costs were smaller than those calculated by Macnamara & Kushnir (1971), although they were not abolished. Furthermore, the more language-switches there were, the slower the overall RT (Experiment 2). That is, a language-switch cost was incurred even when there was no syntactic anomaly[21].

In the light of evidence of cross-language effects (discussed in the next section), Dalrymple-Alford (1985) proposed a different explanation of language-switch costs. He suggested that the two systems were linked to one another. The links, he supposed, were stronger for words of the same language than for words of different languages. Language-switch costs would then arise not because an input-switch had to be re-set (to channel input to the correct lexicon), but because the strength of the links differed. That is, assuming that lexical representations are connected to one another, he proposed stronger *associative connections* between words of the same language than between words of different languages. Dalrymple-Alford (1985) did not further specify the nature of these links or their functioning,

---

[21] See also Meyer & Ruddy (1974), who employed a two-word display paradigm to the same end.

but presumably the notion is one of spreading activation (see Quillian, 1962; Collins & Loftus, 1975; Anderson, 1983; Dell, 1986). That is, the stronger the links between lexical representations, the more easily activation can spread from an activated lexical representation to all other lexical representations connected to it. In the case of strong links, a lexical representation's activation can be raised from its 'resting level' when activation spreads to it. If 'pre-activated' in this manner, it is easier for that representation to be activated above threshold when the corresponding item is presented as a visual stimulus. In the case of weak links, little activation spreads to the associated lexical representations, thus not raising the baseline activation much. Lexical representations of the same language benefit more from spreading activation, because they are linked by stronger associative connections than is the case for lexical representations of different languages, hence, on the associative connections account, there is a cost for switching from one language to another because there are different degrees of 'pre-activation'.

Dalrymple-Alford's (1985) account is a positive development away from the view of two entirely independent lexica, which the input-switch account had espoused. However, the postulation of weak links between the two languages' lexica does not allow for much interaction between the two languages, which is problematic given findings such as de Groot & Nas's (1991) between-language (masked) repetition-priming (if one assumes lexical links to be activatory), or given interlingual word-neighbourhood effects (e.g., van Heuven, Dijkstra & Grainger, 1998; see section 1.5.2.4).

There are three points to note with regard to the input-switch and the associative connections accounts. Firstly, just as Macnamara & Kushnir (1971) had done, Dalrymple-Alford evaluated overall list completion times. That is, these studies do not provide any on-line measure of the dynamics of language processing. Secondly, neither model permits much (if any) interlingual interaction, which is incompatible with many phenomena of interlingual interference, such as accidentally using a German word in an English sentence, but giving it an English pronunciation, thus failing to convey the intended meaning (cf. the example given at the beginning of this chapter). There is also experimental evidence of interaction between the two languages, as shall be described below. Finally, neither the input-switch nor the associative connections account provides a mechanism for controlling the language system. Activating one language over the other, or processing in one language rather than the other, are due to the mode of representation (separate lexica, accessed separately). If, as proposed by Macnamara and colleagues, the languages are separate, how does an individual manage to translate? To read the word CAT and know that the translation equivalent (in German) is KATZE, there needs to be some connection between the two. Dalrymple-Alford's account provides connections between the two languages, which could be used in translating. However, without some control mechanism, how does a bilingual, on the associative connections account, manage to choose the word CAT over the word KATZE when wanting to express the meaning {cat}? (The input-switch account also has this problem.)

As outlined above, Dalrymple-Alford, in contrast to most of the earlier literature (including Macnamara & Kushnir, 1971), suggested that the two lexica were *not* independent of each other, but that they were interconnected, that is, that they were *interdependent*. In short, finding a switch-cost can be accommodated even when postulating interaction between the two language systems. In fact, as Dijkstra & van Heuven state (1998, p. 191; square brackets added): "We must reject the language-selective hypothesis [i.e., selective access to independent lexica] as soon as reliable evidence for on-line cross-language effects are available ...". Is there such evidence?

In language production there is the phenomenon of interlingual Stroop effect[22], which means that for a German/English bilingual the time to name the ink colour red is not only delayed when the letters spell GREEN (as opposed to RED or XXX), but also when the letters spell GRÜN ('green' in German). That is, there is evidence for both intra- and interlingual interference (e.g., Preston & Lambert, 1969; Dyer, 1971; Albert & Obler, 1978; Obler & Albert, 1978; Fang, Tzeng & Alva, 1981; Kiyak, 1982; Dornič & Wirberg, 1983; Mägiste, 1984; Tzelgov, Henik & Leiser, 1990). More relevant to this thesis is interlingual interference in perception. Examples of such phenomena are outlined next.

1.5.2.2 Bilingual flanker tasks

Another approach to the question of independent/interdependent lexica and selective/nonselective access has been the bilingual flanker task. In such a task, a participant is shown a target word flanked (above and below) by two copies of another stimulus,

---

[22] see Stroop (1935) for the origins of this task, using unilinguals.

designated as 'to-be-ignored'. Typically, the to-be-ignored flanker words affect processing of the target word. That is, it is easier to classify the target as belonging to the semantic category X if the flankers are also category X, than if they are category Y. For example, the target TABLE is classified faster as 'furniture' when flanked by CHAIR than when flanked by SKIRT (see Shaffer & LaBerge, 1979, and LaBerge, 1981, for unilingual data). Guttentag, Haith, Goodman & Hauch (1984) extended this paradigm to the bilingual domain. They predicted that, in a bilingual version of the task (e.g., TISCH ['table' in German] flanked by CHAIR vs. SKIRT), such effects would only become apparent if both language systems were active and accessed at once. If, instead, as the input-switch account suggests, only one system at a time were active, then there should be no effect of relatedness between the simultaneously presented flankers and targets (or if there were, RTs should be delayed, given sequential access). In all three experiments, Guttentag et al.'s key finding (for the present purposes), is that even in a cross-language condition (i.e., flankers and targets were of different languages), participants were slower to make semantic category decisions about the target when flankers and targets were of different semantic categories than when they were of the same category. In short, the bilingual participants were affected by the semantic status of the flankers, even when they were in the other language. Furthermore, RTs in cross-language and same-language conditions were comparable. If there is an input-switch, which is re-set in order to process targets and flankers of different languages, then, given that the two languages are processed sequentially (rather than in parallel), there should either have been slower RTs in the cross-language than in the same-language conditions, or the semantic category of the flanker should not have affected RTs (to target words) in the cross-language condition.

1.5.2.3 Masked cross-language priming effects

In Guttentag et al.'s (1984) semantic categorisation task, flanker words were clearly visible. Consequently, responses could have been influenced by various strategies. As de Groot & Nas (1991) suggested, 'post-lexical integration strategies' could take place in such a situation, whereby the individual looks for some relationship between the prime and the target. If such a relationship is found (be this semantic relatedness between the two stimuli, orthographic similarity, rhyme, etc.), then, when required to press button X for the category 'furniture' (e.g., the target is TABLE), there could be a bias towards that response, in the event that the flanker is consistent with the same response (e.g., CHAIR). By the same token, there would be interference to generating response X if the flanker's semantic category is associated with response Y.

In order to investigate cross-language effects, whilst excluding a role for such strategic processes, de Groot & Nas (1991) conducted a bilingual, primed lexical decision task[23], in which they contrasted priming effects in masked and unmasked priming conditions. They argued that masking the primes (which could serve as a bias towards the required 'word' response), would block conscious processing of primes, thus blocking 'post-lexical integration strategies' and retrieval of episodic traces, but not spreading activation. Accordingly, Dutch/English bilinguals' performance was compared in between- and within-language repetition-priming. The most important finding in relation to the question of independent/interdependent lexica, is that there was significant between-language repetition-priming under masked conditions, even when primes and targets shared only their meaning (e.g., BLANKET primed DEKEN ['blanket' in Dutch]). Such a finding suggests interdependent, rather than independent, lexica.

1.5.2.4 Cross-language word-neighbourhood effects

Further evidence refuting the notion that words from the two languages are represented in independent systems, to which access is selective, is found in a study manipulating word neighbourhood. Grainger & Dijkstra (1992) found that French/English bilinguals, performing an English-specific LDT, were slower to respond 'yes' to English words with more French than English word neighbours ('traitors'), than they were to respond 'yes' to English words with more English than French neighbours ('patriots'). Such an outcome is not possible if assuming that the languages are independent of one another. Grainger & Dijkstra (1992) presented these data without providing statistical evidence. Additionally, there was no unilingual control group, which would be able to show that the effects found for bilinguals were indeed effects of bilingual processing, and not artefacts of, for instance, the items used. It is conceivable, for example, that 'traitors' were orthographically less familiar in English than 'patriots', which could have slowed their recognition times.

The above shortcomings were amended in a study by van Heuven, Dijkstra & Grainger (1998), who used lexical decision (LDT) and progressive demasking (PDM) tasks.

---

[23] primed lexical decisions: lexical decisions are made more rapidly if the word has been encountered before in an experimental session (see Scarborough, Cortese & Scarborough, 1977, for unilingual data).

Grainger (1998), who used lexical decision (LDT) and progressive demasking (PDM) tasks. The most pertinent experiment for the present purposes (Experiment 4) presented English words and nonwords to Dutch/English bilinguals and English unilinguals, who were asked to make English lexical decisions. The bilinguals were delayed in their English lexical decisions for 'traitors', but not for 'patriots', while English unilinguals showed no RT difference between 'traitors' and 'patriots'. Experiments 1-3 also showed significant cross-language effects of this type. (See also Bijeljac-Babic, Biardeau & Grainger, 1997, who used a masked primed lexical decision task with brief prime exposures, comparing performance on target words preceded by orthographically similar or dissimilar prime words from the target or non-target language. RTs were slower when primes were orthographically related than when unrelated, both within and between languages.)

Such cross-language neighbourhood effects are in accord with the view that the lexica interact. If the lexica were independent of one another, there should be no effect of the other-language neighbourhood size.

1.5.2.5 Interlingual homograph (IH) interference effects[24]

Much of the extant research directed at the question of independence/interdependence of the lexica has made use of interlingual homographs (IHs). Two types of IHs are distinguished. If both spelling and meaning are shared across the two languages, the item is called a *cognate interlingual homograph* (e.g., for English and German: HAND). If only spelling, but not meaning, are shared, it is called a *noncognate interlingual homograph* (e.g., for English and German: TAG ['day' in German]). The following will show how the interpretation of IH studies (using either type of homograph) has been taken as evidence by proponents of both the independent and interdependent models. Furthermore, it shall be discussed how data previously taken as evidence by proponents of the independent lexica/selective access view, can be convincingly reinterpreted as being consistent with the interdependent lexica/non-selective access view.

Gerard & Scarborough (1989) carried out a 'blocked' lexical decision task with English/Spanish bilinguals. In one block, their lexical decision task required participants to

---

[24] Interlingual homograph (IH) = a letter string existing in both languages (e.g., TAG, meaning 'day' in German). IH interference is investigated by asking bilinguals to make lexical decisions in one language only. However, besides English words and nonwords, the stimulus list includes IHs. IH interference is present if the reaction time to the IH is slower than that to a matched control word (that is a word only in the target language).

decide whether or not letter strings were Spanish words (items were Spanish words and Spanish-like nonwords). In another block participants decided whether letter strings were English words or not (items were English words and English-like nonwords). Included among the real words were non-cognate IHs (same spelling, different meanings), cognate IHs (same spelling, same meaning) and non-homographic translations (different spelling, same meaning). Gerard & Scarborough (1989) found no between-language repetition effect (faster RTs on repetitions) for words that shared their meaning, but not their spelling, across languages (non-homographic translations). Furthermore, they found that the reaction time to non-cognate IHs (e.g., RED ['net' in Spanish]) appeared to be determined by the target language's word frequency, not the non-target frequency. That is, in the Spanish block, lexical decisions to RED (where it has a low word frequency) were slower than in the English block (where it has a high word frequency). (See French & Ohnesorge, 1995, for similar findings with French/English bilinguals.) The absence of repetition priming for the non-homographic translations was interpreted as indicative of separate language systems, where processing a word as a language A word has no effect on the equivalent language B representations. The finding that frequency effects appeared to be determined only by the target language was also interpreted as indicative of independent, selectively accessed lexica: RT was slower when the word's frequency was low in the target language.

However, Gerard & Scarborough's (1989) findings are also in accord with the notion of non-selective access to interdependent lexica. Assume that the presentation of the letter string RED activates both Spanish and English lexical representations, giving rise to 'English word' and 'Spanish word' signals. Given Spanish-specific lexical decision task instructions, 'Spanish word' signals are evidence for a 'yes' response (because the item is indeed a Spanish word). However, at the same time, a representation coded as 'English word' is activated above threshold. The resulting 'English word' signal is evidence for a 'no' response (because instructions are to respond 'yes' only to words existing in Spanish). Thus, it is conceivable that the IHs presented by Gerard & Scarborough (1989) led to competition between the two responses ('yes' / 'no'), observed as a delayed response.

Furthermore, on the basis that the English reading was of higher frequency than the Spanish reading, the 'English word' signal would presumably have been available earlier than the 'Spanish word' signal. Being available earlier, activation of the units underlying a 'no' response could have received activation earlier than the units underlying 'yes' (that relied on the signals from the LF Spanish reading). Receiving activation earlier, the units underlying 'no' would have an activation head-start over the 'yes' response units, making it more difficult to achieve sufficiently higher activation of 'yes' compared to 'no' units, in order to ensure that the ensuing response is 'yes', rather than 'no'. Evidently, Gerard & Scarborough's (1989) findings can equally well be accommodated by the view that lexical access is non-selective and that the two lexica are interdependent, rather than independent.

Dijkstra, van Jaarsfeld & ten Brinke (1998), using Dutch/English bilinguals, performed two studies in which an English-specific LDT ("respond 'yes' if the letter string is a word in English, respond 'no' if it is not") was used to examine the question of interdependent/independent lexica. In this case the frequency of the noncognate IHs' Dutch and English readings were varied orthogonally[25]. In accord with the above discussion, the noncognate IHs that were LF in the target and HF in the nontarget language, were 21ms slower than the matched English control words. However, this difference was not significant, which could be regarded as evidence that there is no cross-language interaction. However, the variance for these items was relatively large, which may have prevented a significant effect from being demonstrated. A further reason for the null-finding may be the list type. That is, since no purely Dutch words (that would have required a 'no' response) were included, the participants may have biased their decision criterion towards 'wordness' signals. That is, information such as familiarity or high levels of general activation within the lexicon, may have been used to calculate (while the actual processing of the item was still taking place), whether the letter string in question was more likely to be a word or a nonword, biasing either towards a 'yes' or a 'no' response[26]. In the case described above, if such wordness signals were allowed to play a significant role (permissible only if no pure non-target language words are presented, for which there would be the danger of falsely accepting them), then making use of wordness signals could attenuate the adverse effects produced by the HF non-target language reading activating 'no'. This is because the wordness signals will feed activation into the 'yes' response units, helping them to achieve higher activation than the 'no' response units.

In sum, evidence from IHs previously taken as evidence for selective access to independent lexica, can be reinterpreted in favour of non-selective lexical access to

---

[25] LF in both languages; HF in both languages; LF in Dutch /HF in English; HF in Dutch / LF in English.

[26] This is referred to as 'paralexical decision processes' (Monsell, Doyle & Haggard, 1989; cf. also Grainger & Jacobs, 1996, who use the term 'multiple read-out'), and shall be described in more detail later on.

interdependent lexica. Furthermore, null-effects (e.g., Dijkstra, van Jaarsfeld & ten Brinke, 1998, see above) can also be accommodated by this latter view. What has not been addressed by the IH literature so far, is how, during the processing of such items, control over the two lexica is effected. This is the topic of Chapter 5, where the IH literature will be discussed in further detail.

As this review has shown, much valuable progress has been made with respect to the cognitive representation of a bilingual's two languages. Starting from the early position where it was unclear whether there are language-specific representations, or whether language is stored in an abstract way (see Kolers, 1968), there is now convincing evidence that lexical representations have language-specific representations, while semantic representations do not (e.g., Durgunoğlu & Roediger, 1987). However, none of the above studies on, or models of, bilingual language processing addressed the question of how a bilingual's two languages are controlled. This is particularly important given the evidence that the two languages' representations interact.

As has been outlined here, there is convincing evidence that the two lexica are inter- rather than independent and that access to them is non-selective rather than selective (e.g., IH interference). At the same time, there is convincing experimental evidence that switching languages can incur a cost. The coexistence of these two phenomena is inexplicable with recourse simply to the way in which the two languages are represented, since the two phenomena are then contradictory. That is, if the languages are represented in a way so as to have parallel access to interactive lexica, there should be IH interference, but no switch-costs. In contrast, if the languages are represented in a way so as to have selective access to independent lexica, one may expect switch-costs, but not IH interference. Instead of this either/or view, which arises when focusing solely on how the languages are represented, an approach is required which regards the language processing system as an interactive, dynamic system.

The question of control appears to have been addressed only by two models, The Bilingual Interactive Activation (BIA) model (e.g., Dijkstra & van Heuven, 1998) and the Inhibitory Control (IC) model (e.g., Green, 1998a, b). These two are the key models evaluated in this thesis.

1.6 The two 'control' models: Bilingual Interactive Activation / Inhibitory Control

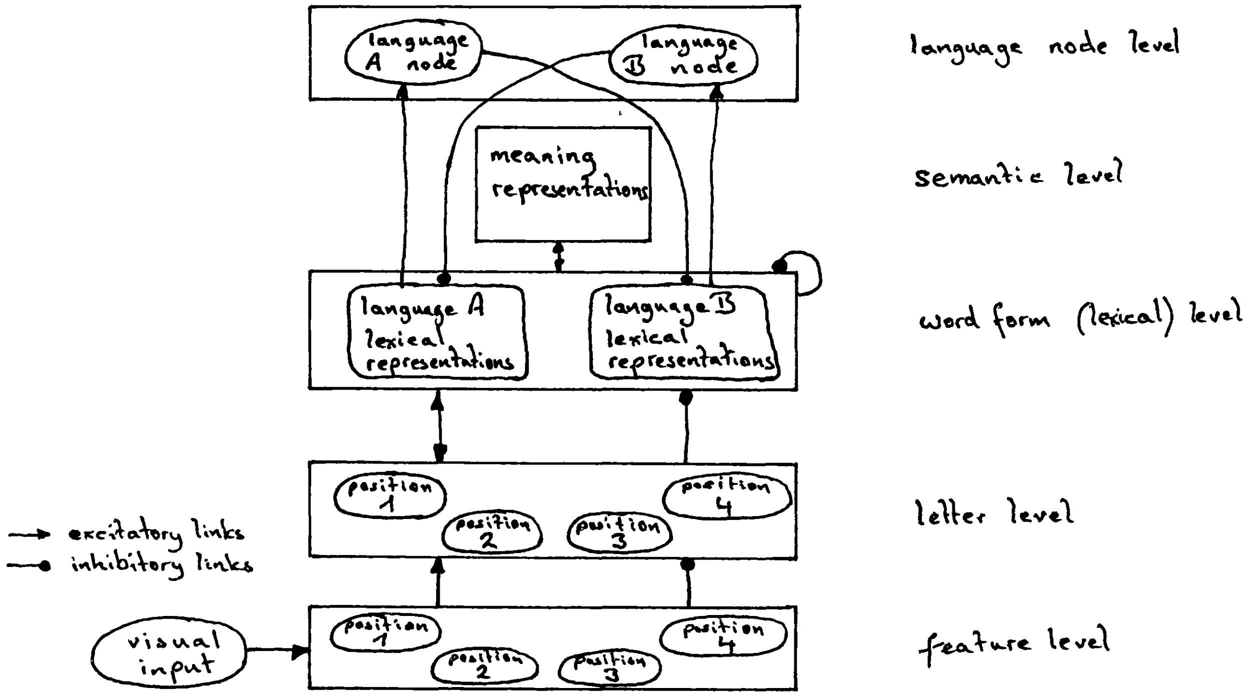
The BIA and IC models provide a means of accommodating the assumptions of non-selective access to interdependent lexica while at the same time accounting for the phenomenon of language-switch costs, that had previously been taken as the behavioural evidence of selective access to independent lexica. The following presents the two accounts, and assesses whether they are able to accommodate currently known phenomena of bilingual language processing.

1.6.1 An internal account of control: The Bilingual Interactive Activation model

Dijkstra, van Heuven & Grainger (1998; see also Grainger & Dijkstra, 1992; Dijkstra & van Heuven, 1998; van Heuven, Dijkstra & Grainger, 1998) developed a computational model of visual bilingual word comprehension. This model is a bilingual version of McClelland & Rumelhart's (1981) Interactive Activation model of unilingual visual word recognition. In the BIA model, several representational units (for features, letters, words, languages) are organised in a hierarchical network. The bilingual model differs from the unilingual version in two respects. Firstly, it has lexical representations for two languages. Whether there are two fully equipped lexica, or whether some representations are shared, is at present unresolved. Specifically, words that have the same spelling in both languages (IHs) could be represented once, shared between the languages, or they could be represented twice, once for each language. Considering language-specific frequency effects (Dijkstra, van Jaarsfeld & ten Brinke, 1998; de Groot, Delmaar & Lupker, 2000; and Dijkstra, Timmermanns & Schriefers, 2000), Dijkstra, Grainger & van Heuven (1999) suggest that the separate-representations version is perhaps the more accurate model. The second difference between the BIA and the IA model is that language nodes were introduced. This permitted the language affiliation of each lexical representation to be characterised, as well as allowing the overall activation of a language to be affected by the presentation of just a single word (cf. Grosjean, 2001, who speculates that a single word in another language might be sufficient to shift individuals towards the bilingual end of the so-called unilingual-bilingual continuum). A semantic level was later added into the model by Dijkstra, van Jaarsfeld & ten Brinke (1998). As this thesis utilises both lexical and semantic tasks, this is the version depicted in Figure 1.2.

Figure 1.2

Dijkstra, van Jaarsfeld & ten Brinke's (1998) Bilingual Interactive Activation model (but see also Dijkstra, van Heuven & Grainger, 1998).



The model functions as follows. Upon perceiving a letter string, the features within the letter string are analysed and corresponding feature representations are activated. Then, the corresponding letter units are activated, which in turn activate the lexical representations with matching letters in the correct positions (as described earlier for the IA model). At this stage, both the lexical representation of the presented stimulus itself, and also, to a lesser extent (because there is less of a match), its lexical neighbours, will be activated. Active lexical representations feed activation to the language-node to which they are connected. An activated language-node inhibits lexical representations of the other language. Active lexical representations compete with other active lexical representations by means of lateral inhibition. This capability allows within- and between-language effects to be simulated, such as the 'traitor' and 'patriot' effects reported by Grainger & Dijkstra (1992), described above. Apart from the 'bottom-up' activation described so far, the BIA model also allows for 'top-down' activation. For example, if the lexical representation CAT is active, it sends activation to the letter representations C, A and T. Given a context (as when reading a text or a sentence), there may be expectation of what the next word is likely to be (see discussion in Pinker, 1994). In that case the representation most compatible with that expectation may receive activation top-down, even when that word has not yet been encountered.

Using bilingual between-language repetition priming in an LDT, de Groot & Nas (1991) found that participants were faster to respond 'yes' to the target 'DEKEN' ('blanket' in Dutch) if primed by 'BLANKET', even when the prime was masked. Such non-homographic cognates do not share their spelling, and so must have individual lexical representations. As described above, masking the prime, de Groot & Nas (1991) argued, blocked post-lexical integration processes and retrieval of episodic trances, but not spreading activation. These authors suggested that activation would spread from one lexical representation (BLANKET) to the other (DEKEN) via lexical links, pre-activating the target word's lexical representation, and thus shortening activation time and hence lexical decision time. That is, on their account lexical links convey activation; according to the BIA model, lexical links are inhibitory. How can the two be reconciled?

One way of thinking about de Groot & Nas's (1991) between-language repetition-priming effect is to consider connections to semantic representations. In the cross-language repetition priming context, the language A word (e.g., DOG) has a lexical link with its translation equivalent (HUND, 'dog' in German). According to the BIA model, these links are inhibitory, but the effectiveness of the inhibition they can exert may be overridden by positive feedback from the semantic level. In brief, given the prevailing evidence that semantic representations are shared between languages (and that activatory links between lexical and semantic representations are bi-directional, see Figure 1.2), one could assume the following. Having activated the lexical representation DOG, the corresponding semantic representation {dog} is activated, which feeds activation back down to the lexical representation HUND (as well as DOG), raising its activation level (cf. Grainger & Frenck-Mestre, 1998). Lateral inhibition should lower the activation level of HUND. Given the robust repetition-priming effects found in de Groot & Nas's (1991) masked primed LDT, one may assume that if lexical links are indeed inhibitory, top-down semantic activation is overall stronger than lateral lexical inhibition, since otherwise HUND should have been delayed by the preceding DOG. That is, there should have been between-language inhibition of translation equivalents, rather than priming. (See also Gollan, Forster & Frost, 1997, for masked cross-language translation priming between different scripts.)

Another simple effect generally found in bilingual experiments is that the bilingual participants respond faster to L1 than to L2 words (e.g., Thomas & Allport, 2000). This imbalance is implemented in the BIA model by assuming lower subjective frequency of L2 than of L1 words, i.e., lower resting activation levels for L2 than for L1 lexical representations. Presumably the more balanced the bilingual, the more comparable the resting activation levels of both languages' word representations. In that case RTs for the two languages would be expected to be comparable (e.g., Soares & Grosjean, 1984).

The relative activity of the two lexica is modulated top-down, by inhibition from the language-nodes. That is, during L1 processing, the L1 language-node inhibits L2 lexical representations. The stronger this top-down inhibition, the less lateral-inhibition should be able to influence the target language. This, in fact, has been used as the explanation for the absence of IH interference effects by Dijkstra, van Jaarsfeld & ten Brinke (1998, Experiment 1), since lateral-inhibition between IHs' lexical representations is assumed to lead to an RT delay on such items. The stronger the top-down inhibition, the weaker the lateral-inhibition, and consequently the weaker the IH interference effects observed. In Chapter 5, this is discussed in more detail, where Experiment 4 contrasts the BIA and IC models, testing them against IH interference data.

In the BIA model, the phenomenon of language-switch costs is also explained by the notion of top-down inhibition. Processing in language A entails inhibition of language B representations. Thus, when switching to language B, a decrement in activation (simply 'activation decrement' in the following) needs to be overcome. Therefore, whenever there is a language-switch, there should be an RT delay. This concept is important in particular with regard to Experiment 3 (Chapter 4), which employed a semantic categorisation task. In the BIA model, semantic representations are not subject to the top-down inhibition from the language-node (see Figure 1.2). However, accessing semantic representations involves activating lexical representations. Since lexical representations are subject to top-down inhibition, the BIA model predicts a language-switch cost even when a task targets the non-suppressed semantic level.

One disadvantage of the BIA model is that it does not clearly specify how signals from the system are mapped onto responses, or what signals from the lexico-semantic system are used. It is possible that, in a lexical decision task, a 'yes' response is given once one language-node has achieved dominance. However, this could lead to false positives on 'wrong language' items (e.g., in a language-specific LDT, a German word presented on an English trial). To avoid such errors, there is a need to incorporate a verification mechanism that tests the output from the lexico-semantic system ('German word') against the goal ("reply 'yes' only if it is a word in English") before a response is allowed to be activated above threshold. To date, such a mechanism has not been incorporated into the BIA model. Secondly, it is unclear how the relevant signals are linked to the response units. Apart from any empirical evidence that the experiments here might show, the Inhibitory Control model has an advantage over the BIA model in that it specifically addresses the question of what signals may influence responses and, most importantly, how signals are mapped onto responses.

1.6.2 An external account of control: The Inhibitory Control model

Earlier in this chapter, a reference was made to a 'linking mechanism', which links signals from the lexico-semantic systems to the response units. The IC model provides such a mechanism. This model is an adaptation and development of Norman & Shallice's (1986) and Shallice's (1988; 1994) task schema model for the control of action. Given the importance of the notion of task schemata for the IC model, a brief overview of the meaning of 'task schema' is provided first.

1.6.2.1 Task schemata

A task schema may be roughly defined as the set of cognitive operations required to perform a task effectively. Throughout this thesis, the notion of 'task schema' is used as the language equivalent to the schema notion developed by Norman & Shallice (1986) for their model of action. To perform any action (physical or communicative), an individual needs to be able to define this action as a goal to be achieved. When there are two different actions to be switched to and from, then two goals need to be specified, and the individual must ensure that the correct goal is pursued at the correct time. For example, when wanting to speak in German, the task schema that needs to dominate could be defined broadly as a 'speak German' task schema. When the same individual then wants to express something in English, the task schema 'speak German' will no longer be correct. Instead, the individual will need to ensure that a 'speak English' task schema dominates over all other existing task schemata, but particularly over the previously dominant 'speak German' task schema.

Norman & Shallice's (1986) model of action incorporates a complex interaction between exogenous (external) and endogenous (internal) control. It proposes a structure of controlling mechanisms, organised in a hierarchical structure, working in concert and influencing each other. Components of a familiar task are organised by low level schemata, which are activated by appropriate input. The higher level 'contention scheduling' (CS) mediates competition between schemata that are triggered by the same input. To ensure that infrequently used schemata can be selected, as opposed to frequently used ones, a superordinate layer of control was proposed, the 'supervisory attentional system' (SAS), which determines goals (e.g., make lexical decisions about words, rather than simply reading them) and establishes task schemata for novel tasks in response to instructions (e.g., a 'lexical decision task schema'). In short, higher levels of control modulate the lower level schemata that actually organise the action, but external stimuli can also trigger schemata of automated cognitive processes (e.g., to read a word rather than name its ink colour, as in the Stroop task). The following outlines how the theory of task schemata may be envisaged in practice, in particular in the context of lexical decision and animacy decision tasks (the two types of task employed in this thesis).

If schemata can be activated through exogenous influences, then one implication, for example, is that, in a lexical decision task (LDT), seeing a German word may trigger a task schema for making a lexical decision in German. The necessity of endogenous control becomes clear here, for example to maintain the goal of making lexical decisions in English, not in German (should German words be presented either on trials designated as English or in an English-specific LDT). Once a schema's goal has been attained, it is deactivated (e.g., having completed a lexical decision task and leaving the laboratory, the individual does not continue to make lexical decisions about the words he perceives). A schema may be deactivated directly by the SAS if another goal is required, or by being superseded by another schema with a higher activation level that was triggered bottom-up. Competing schemata are in a mutually inhibitory relationship (i.e., each can inhibit the other). For example, in a German/English language-specific LDT, there is mutual inhibition between the German-decision task schema and the English-decision task schema[27].

Having outlined the notion of task schemata that underlies the IC model, this external account

---

[27] Note that the capabilities attributed to the SAS in the original versions of Norman and Shallice's model (e.g., Norman & Shallice, 1986, Shallice, 1988), are now regarded as too overarching. Increasingly, it is thought necessary to fractionate such a strong central executive (see Monsell & Driver, 2000). This thesis makes use of the notion of task schemata, as proposed by Norman and Shallice. However, the work presented here does not directly inform the work related to central executive function. Therefore, the line of investigation and argumentation regarding the powers of the higher levels of control shall not be pursued here.

of control can now be presented, addressing the question of control during language perception.

On the external account, a language-switch cost in perception tasks (e.g., LDTs) is thought to arise predominantly at the output from the lexico-semantic system, where task schemata act differentially to the signals emerging from the lexico-semantic system (see Green, 1998a). The IC model proposes control predominantly at the task schema level (mutual inhibition of task schemata), outside the lexico-semantic system. Green (1998a) contemplates the possibility of inhibition of non-target lemmas (i.e., lemmas with the 'wrong' language tag) in language production, but considers external control (i.e., task schema level) to be critical for perception tasks. In any event, this model does not suggest inhibition at the lexical level (as the BIA model does). How does external control work?

Given LDT instructions, a lexical decision task schema is set up. This schema links signals indicating that the item is a word (e.g., that a lexical representation was identified) to the units underlying a 'yes' response. Signals indicating the presence of a nonword (e.g., that there was no lexical representation), are linked to the units underlying a 'no' response. For bilinguals, words are words of either one language or another, providing not just 'wordness' information but also language information (e.g., for the stimulus CAT: "it is a word, it is English"). The language membership of a word is likely to affect the bilinguals' processing in a language task such as lexical decision. Specifically, one may assume that the architecture of task schemata is influenced by the instructions given; this is discussed in the next section.

1.6.2.2 Task instructions and task schema 'architecture'

This section outlines the task schemata thought to be operating in the tasks employed in this thesis (lexical decision and animacy decision tasks).

Language-specific LDT instructions[28] encourage the establishment of two task schemata, one to link "German word" signals to the 'yes' units, the other to link "English word" signals to the 'yes' units. Given that only one schema at a time can drive a response, one schema must dominate. This can be achieved by the two schemata being in mutual inhibition (the selected schema inhibits the non-selected one). Thus, for example, on a German-decision trial, the German task schema needs to dominate over the English one (see

---

[28] e.g., "decide whether the letter string is a word in German when the computer screen is yellow; when it is blue, decide whether the letter string is a word in English"

Figure 1.3). Task cues, such as screen colour, can serve as a signal that a given schema is required. For example, given instructions to make a German lexical decision when the screen colour is yellow, perceiving a yellow screen could trigger the German LD task schema (in the event of a language-switch trial; on a language-non-switch trial task cues will simply reinforce the already dominant schema). Additionally, signals from the item itself, indicating that it belongs to a given language, may also perform such a function. This could be information available early on in processing the visual input (such as orthographic information, e.g., the letter combination PF, which is permissible in German but not in English). Language information would also become available later, once the item has been processed in sufficient detail to have assessed its language affiliation, e.g., by having obtained its language tag (e.g., Albert & Obler, 1978; Green 1986), or by having established which language-node it connects to, or once one language node has become dominant (Dijkstra & van Heuven, 1998).

In a language-specific task, an arrangement of strong mutual inhibition is necessary so as to be able to accept only German words on German trials and successfully reject any English words should they appear on German-designated trials (called 'wrong language words' in the following). If there were insufficient inhibition, then a wrong language word could trigger the corresponding task schema. Since an LD task schema links word signals to the units underlying a 'yes' response, the consequence would be a false positive reply (e.g., replying 'yes' to an English word on a German trial).

In a language-specific LDT a change of target language requires a change of task schema. When making a decision in the same language as on the preceding trial, the same task schema, as before, can be used again. Since the reconfiguration process takes time[29], RTs on language-switch trials are slower than those on non-switch trials; i.e., there is a switch-cost[30]. The stronger the mutual inhibition, the larger the language-switch cost.

---

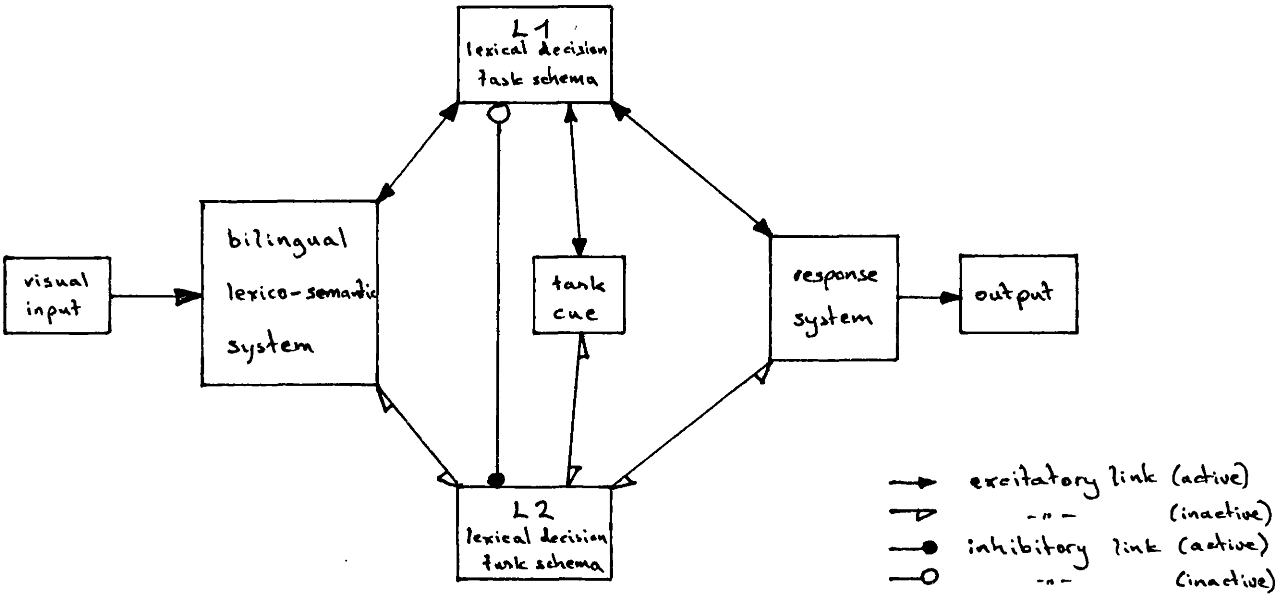
[29] Whether this cost reflects the time required by the control processes to reconfigure the cognitive system ('task-set reconfiguration' e.g., Monsell, 1996), or whether it reflects competition between task schemata, characterised by persisting activation and inhibition ('task set inertia'; Allport, Styles & Hsieh, 1994), is the subject of debate, but shall not be pursued here (for a discussion of the two positions see Monsell 1996).

[30] This is a temporal switch-cost. There may also be a cost to accuracy: If the participant favours speed of response over accuracy, he may decide (having just switched from an English-decision trial to a German-decision trial) "no, the item is not an English word", when the letter string is a real German word, and so should have been responded to with 'yes'. The extent to which a temporal or an accuracy cost arises, is determined by the speed/accuracy trade-off adopted by the participant.

Figure 1.3

An Inhibitory Control model (after Green, 1998a).

language-specific lexical decision task (eg. Experiment 1a), L1 trial:



To fully reconfigure a task schema set up, it appears that the input stimulus is required, rather than this process being fully subject to (top-down) voluntary control (see Allport, Styles & Hsieh, 1994; Rogers & Monsell, 1995; Meiran, 1996; Meuter & Allport, 1999). For example, Rogers & Monsell (1995) found switch-costs even when participants had ample time to prepare for a switch. Even when participants know which language is required on the next trial (e.g., in a regular-switching paradigm), they apparently cannot fully reconfigure the system in anticipation.

Given instructions to "accept if it is a word, either in German or in English" (a 'language-general' LDT) individuals could establish two separate task schemata, but allow the two to be jointly active. In this case, a given schema could become temporarily more active when a relevant language signal is detected. Assuming that task sets' activation levels do not immediately revert to zero after use, there would be a benefit of having just used a given task schema. In short, one could expect a language-switch cost for a language-general LDT, just as for the language-specific LDT. But without mutual inhibition between the schemata, it should be smaller[31], since, by definition, there is no possibility of 'wrong

---

[31] Even if one assumes mutual inhibition between task schemata in a language-general LDT, then one may suppose that this inhibition is weak, reflecting the relaxed task instructions ("respond 'yes' if it is a word in either language"). That is, the switch-cost should be smaller than in the language-specific LDT.

language' words in a language-general LDT, and therefore no *need* to inhibit the other task schema.

Theoretically, it would be possible to perform a language-general LDT without reference to language. In this case there could simply be a single task schema: "link 'word' signals to 'yes', link 'nonword' signals to 'no' ". If bilinguals were able to act solely on the 'wordness' information of a letter string, and if there were only one task schema, then one might argue that there should be no language-switch cost at all in the language-general LDT. However, in a task that targets lexical representations (at which level there is a distinction between the two languages), it is considered unlikely that bilinguals can ignore the fact that a word is a word in a particular language (see also Kroll & Dijkstra, 2001). On the assumption that bilinguals associate different cognitive computational processes with each language at this level of representation[32], it is thought that they are unlikely to establish a single task schema even when the task instructions would permit it, but instead establish a task schema for each language, (cf. Thomas & Allport, 2000). In short, one may expect a language-switch cost in a language-general LDT, but it should be smaller than in a language-specific LDT. If bilingual participants *did* establish only a single task schema, then there might still be a language-switch cost in a language-general LDT: Language signals, it was argued, can presumably not be ignored. Perceiving a change of language signal could bias towards a change of response, as discussed in section 1.6.4.

In contrast, in a semantic task, there is a greater likelihood of bilingual participants being able to establish a single task schema (given language-general task instructions), since language-neutral semantic representations presumably do not normally have separate processes associated with them. This would suggest that there may be a situation where there is a switch of language but no language-switch cost. However, language membership information will presumably become available during word processing, and it is thought

---

[32] Lexical representations differ for different languages, reflecting the respective word-form (e.g., MOON vs. MOND vs. LUNE ['moon' in German and French respectively for the latter two]). These language-specific word forms require different syntactic rules, quite apart from different phonological rules. For instance, MOND has a male article (der), while LUNE has a female article (la), with implications for how they are conjugated. For instance, in German, articles (and in some cases word endings) change with case, but differently depending on the gender (e.g., male: der Mond / des Mondes / dem Mond / den Mond; female: die Frau / der Frau / der Frau / die Frau; neutral: das Gesicht / des Gesichtes / dem Gesicht / das Gesicht [nominative, genitive, dative and accusative, respectively]). Adjectives remain unchanged by the gender of the noun they refer to (e.g., der schöne Mond / die schöne Frau / das schöne Gesicht [the beautiful woman / moon / face, respectively, in German]). In French, in contrast, the article remains the same, regardless of case, but adjectives vary, depending on the gender of the noun they refer to (e.g., la belle lune / le beau visage [the beautiful moon / face, respectively, in French]).

unlikely that a bilingual participant is able to ignore such language membership information. Language information could influence the participant's decision 'paralexically', as described in section 1.6.4.

1.6.3 List composition effects

As indicated in the above, task schemata and their usage are not the sole determinants of processing in lexical decision tasks. In the literature, there is evidence that nonword type can affect response criteria, modulating frequency effects (i.e., the more word-like the nonwords in a stimulus item list, the clearer the difference between HF and LF RTs; see Monsell, Doyle & Haggard, 1989; Stone & van Orden, 1993). The explanation for this effect is that, depending on list composition, a more relaxed or a more stringent response criterion is used.

A relaxed criterion means using global activation levels (and familiarity information[33]). For example, perceiving that there is a high level of activation in the lexicon indicates the presence of a word. A stringent criterion means processing items more thoroughly (e.g., to the point of having uniquely identified the stimulus). A stringent criterion needs to be used when, for example, HF and LF words are accompanied by word-like nonwords (e.g., SWERM), because the amount of general activation in the lexico-semantic system may be very similar for the LF words and the word-like nonwords. In this case, RTs to HF words are clearly faster than those to LF words. In contrast, when HF and LF words are accompanied by non-words that are unlike real words (e.g., SGAJK), it is possible to use a relaxed response criterion, since nonwords will not give rise to much activation. In this case, RTs to HF words and RTs to LF words are less distinct.

When word-like nonwords are accompanied by only HF words, a relaxed criterion should also suffice, because in this case general activation should be clearly higher for word stimuli than for nonword stimuli, making it a reliable distinguishing criterion. Indeed, Glanzer & Ehrenreich (1979) and Gordon (1983) found 'frequency blocking' effects for unilinguals, using lexical decision tasks: 'Pure HF' and 'mixed frequency' stimulus list conditions were contrasted ('pure': comprising HF words and word-like nonwords; 'mixed': comprising HF words, LF words and word-like nonwords). RTs to HF words presented in pure lists were significantly faster than RTs to the same HF words presented in mixed lists. Both lists used

---

[33] The more familiar a letter string, the stronger the indication that it must be a word (e.g., BUTTER vs. SYRUP).

the same nonwords. That is, given a pure-HF list, individuals appear to be able to adopt a relaxed response criterion, whereas faced with a mixed list individuals adopt a more stringent criterion. A decision criterion may be subject to local dynamic adjustments to some degree (see Gordon, 1983), but overall, once adopted, it appears to be applied throughout[34].

1.6.4 Multiple signals available for use: The paralexical decision process

As described by Monsell, Doyle & Haggard (1989; see also Grainger & Jacobs, 1996), many signals arise during language processing. These may be signals pertaining to the presented items (e.g., wordness signals, language signals), task signals (e.g., signals about the computer's screen colour, or about the current item's location on the screen), and also, presumably, 'change' signals, i.e., whether there is a difference between the current trial and the preceding one (e.g., colour has changed, location has changed, language has changed). These, and other, signals may influence the processing of the task. (Note that data available to date suggest that while multiple signals are used during lexical decision, orthographic representations play a major role; see Pexman & Lupker, 1999.) Monsell, Doyle & Haggard (1989) suggest that such information may also be used to evaluate which response is most likely to be correct for the letter string currently being processed (be this in a lexical decision or in an animacy decision task). This 'paralexical decision process' is important for the experiments presented in this thesis, in particular for Experiments 2 and 3. The relevance of these processes for Experiment 2 (LDT) lies predominantly in language-general LDT experience and list composition effects (described above). For Experiment 3 (animacy decision task: ADT), paralexical decision processes are relevant from the point of view of the influence language signals may have, over and above the animacy signals relevant for the task ('multiple read-out', as Grainger & Jacobs, 1996, call it).

Even when a strict response criterion is adopted, information of the type used for relaxed criteria (i.e., that a letter string is familiar, or that there is a high [or rapidly rising] level of global activation), is conceivably available, and may influence the response.

Wordness signals. In an LDT, the paralexical decision processes could be envisaged

---

[34] It is conceivable, however, that a response criterion changes if the nature of the stimuli changes in the course of a list in such a way as to be detrimental to performance. E.g., if an LDT initially presented only nonwords that are unlike real words and then began to show many word-like nonwords, an initially relaxed criterion would probably become more stringent when errors occurred (and were noticed).

as 'wordness' evidence[35] activating the units underlying the 'yes' response. That is, due to this process, the activation levels of the units underlying the response for which there is most evidence, are higher. Thus, there is a bias towards a certain response. This could effectively speed up reaction time, since once the necessary information is available (high levels of general activation in the lexicon / unique identification of a lexical representation), a response can be given almost immediately. For example, given the letter string 'punt', the individual tries to establish whether he has a mental representation for such a letter string. At the same time, neighbours of this word (hunt, pint, etc.) become active along with the representation of 'punt'. This general activation indicates that there is a good chance that 'punt' is a word, rather than a nonword. On this basis, the units underlying a 'yes' response receive activation. Once the lexical representation for 'punt' has achieved threshold, an appropriate signal can be linked to the 'yes' response units. If the 'yes' response units have received activation paralexically, then it will not take much additional activation (and time) to achieve sufficient activation of the 'yes' units over and above the activation level of the 'no' units.

'Language-change' signals. Perceiving that the language of input has changed, appears to bias the bilingual participant towards the response *not* used on the preceding trial. Thomas & Allport (2000) found that participants displayed the normal language-switch cost when repeating responses. However, when *changing* responses, the switch-cost disappeared - in fact, RTs were now *faster* on language-switch than on language-non-switch trials. That is, response type modified the language-switch cost. (See Rogers & Monsell, 1995, for corresponding data using task-switching and response-switching.) Given two consecutive word trials, participants should give two consecutive 'yes' responses. If the change of language input on a language-switch trial causes a bias in favour of a 'no' response, then, to avoid an erroneous response, the participant would need to activate the 'yes' units sufficiently above the pre-activated 'no' units. This may be assumed to take time, evidenced as a delayed response. However, where the bias towards the previously not used response is correct (e.g., nonword ... word), this bias appears to be advantageous: RTs on language-switch trials were faster than RTs on language non-switch trials when the response changed (Thomas & Allport, 2000).

---

[35] E.g., high or rapidly rising activation in the lexicon.

1.6.5 Competition between response units

To be able to respond 'yes', one might argue that the 'yes' units need to exceed the activation level of the 'no' units. This could be achieved by mutual inhibition, as proposed for the task schemata. If both 'yes' and 'no' units were to start from a comparable level of activation, then the time needed to be able to respond 'yes' (or 'no', for that matter) might be x. If, through paralexical decision processes, 'yes' units have already received y amount of activation, then the time required to be able to respond 'yes' will no longer be x, but be reduced to x-y. In other words, if the correct unit's activation level was boosted, less additional activation is required for it to dominate, than if it was not previously boosted. Should the *wrong* response units receive activation (e.g., if a nonword is very word-like there may be activation of the 'yes' units), then the time required to respond would increase (e.g., to x+z), because it then becomes necessary for the correct response units to achieve sufficiently higher activation than the incorrectly activated response units. This all, of course, assumes that the correct reply is given and erroneous activation of one response's units are recognised as being incorrect. If it is not recognised that the wrong response units were activated, and this activation is allowed to proceed, the wrong response will be given.

Levelt, Roelofs & Meyer (1999) propose 'node' selection on the basis of competition without inhibition. The mechanism described by Levelt et al. does not refer precisely to response units - under consideration here - but more generally to nodes within the lexico-semantic system. Activation spreads between nodes and decays over time. The selection mechanism weighs the activation of the response candidates, contrasting their activation patterns, thus yielding a competition sensitive response mechanism. In this manner a node with more evidence in its favour (more activation) could be identified and chosen. Note that this thesis does not aim to differentiate between an inhibitory selection procedure and a selection procedure based on sensitivity to competition; it simply acknowledges that some processes must occur which allow competition between units to be resolved.

1.6.6 Degree of usage of the paralexical decision information

Recall the 'frequency blocking' experiments (Glanzer & Ehrenreich, 1979; Gordon, 1983) outlined above, in which it was found that a 'pure' list appears to allow participants to employ a relaxed response criterion, in contrast to a 'mixed' list, which appears to require a more stringent criterion. Is it conceivable that, apart from the response criterion adopted, individuals are less or more influenced by the information available via paralexical decision processes depending on their list composition? It seems plausible that this could be so. Consider a frequency blocking scenario, where individuals with a pure list adopt a more relaxed criterion, and individuals with a mixed list adopt a more stringent response criterion. Given a pure stimulus list, individuals experience many instances where the information available via the paralexical decision process has provided the correct response. Given a mixed stimulus list, individuals would experience fewer cases where the paralexical decision process was fruitful. Accordingly, one might argue that, apart from adopting either more stringent or more relaxed response criteria (dependent on list composition), the effectiveness of a paralexical decision process (which is taking place in the 'background', irrespective of response criterion), may also be determined by list composition. That is, if experience shows the signals provided by the paralexical decision process to be correct in a significant number of cases (e.g., in a 'pure HF' condition), more weight may be attached to this information. Should it prove helpful in a restricted number of cases (e.g., in a 'mixed' HF/LF condition), it is likely that less weight will be attached to the information it provides.

In sum, confronted with a list of mixed frequencies (and word-like nonwords), individuals may not only adopt a more stringent response criterion, but may also attach less weight to the 'wordness' signals that also become available 'paralexically'. In contrast, individuals confronted with a list containing only HF words (and word-like nonwords) may adopt a more relaxed criterion. Additionally, they may also attach more weight to 'wordness' signals available 'paralexically'. This is of particular relevance to Experiment 2 (Chapter 3).

1.6.7 Asymmetric language-switch costs

There are various instances, not only in task switching studies, but also in language-switching studies, of a switch-cost being greater for switching in one direction than for switching in the other (for a thorough review of task-switching studies, see Monsell, Yeung & Azuma, 2000). These are called 'asymmetric switch-costs'. How would the task schema account accommodate them? In brief, if task schemata for L1 are more practised than those for L2 (because the bilingual's L1 is dominant), then it is conceivable that the dominant language's task schemata are more easily triggered, since the language processing system is more finely tuned to the signals of that language. That is, for a German-dominant German/English bilingual, it is possible that incoming German stimuli trigger German task schemata more easily than incoming English stimuli can trigger English task schemata. This would be a 'regular asymmetric switch-cost'. There is, however, also the frequently occurring phenomenon of 'paradoxical asymmetric switch-costs', where the cost is greater for switching into the normally *stronger* task or language (e.g., Allport, Styles & Hsieh, 1994; Thomas & Allport, 2000, Experiment 2, personal communication). Such a pattern could arise on the basis that the individual attempts to avoid having their stronger language's task schema being triggered by incoming stimuli. Individuals could achieve this by inhibiting their dominant language's task schema more strongly than their weaker language's task schema. That is, it would be more difficult to 'undo' the inhibition on the dominant language's task schema, entailing greater switch-costs when switching into the stronger language[36]. Note, this thesis did not aim to investigate asymmetric switch-costs, but discusses them where they arise. For the present purposes, the main point is that the IC model can accommodate these phenomena.

1.7 Rationale for the methodology employed

To investigate the question of bilinguals' control over their two languages, various factors need to be considered:

1) Both languages need to be presented to participants. This may be done directly or indirectly, as described below.

2) Different levels of processing need to be contemplated, since they appear to differ with respect to being coded for language or not (lexical vs. semantic representations). Representations coded for language are, on the internal account, under the influence of top-down inhibition, while language-neutral representations are not (see Figure 1.2). In contrast, the external account proposes that neither type of representation is subject to top-down inhibition.

These factors, and the paradigms chosen here which accommodate them, are discussed next.

Presenting both languages to the participant. This may be achieved either directly, by presenting words of both languages (as in a bilingual lexical decision task), or indirectly, in a

---

[36] The BIA model achieves both asymmetric and 'paradoxical asymmetric' switch-costs by the degree of inhibition from the language-node to the lexical representations. Asymmetric switch costs could arise because the stronger language's language node becomes active more rapidly (e.g., because there are stronger signals from the corresponding lexicon, or because the language node is more sensitive to such signals). Paradoxical asymmetric costs could arise because top-down inhibition (node level to lexical level) is set to be stronger for suppression of the stronger language (see Dijkstra & van Heuven, 1998: 'asymmetric inhibition').

'unilingual' LDT, by presenting words that have representations in both languages (e.g., non-cognate interlingual homographs, such as the letter string TAG ['day' in German]). Both methods are valuable, and both were employed[37].

Different levels of representation. Research has suggested that while lexical representations are coded for language, semantic representations are not (e.g., Durgunoğlu & Roediger, 1987; Smith, 1991; de Groot, 1993; Kroll, 1993; Kroll & Stewart, 1994). This is an important factor when contrasting the two accounts of control.

On the internal account, non-target language lexical representations can be inhibited by the target language's language-node, but the language-neutral semantic representations are not subject to this inhibition (see Figure 1.2). Any task requiring activation of representations that are subject to this inhibition, should, when the target language changes, show the effects of previous inhibition. That is, both a task that targets lexical representations, and also one that targets semantic representations, should show the effects of lexical representations having been suppressed (i.e., there should be language-switch costs both in lexical decision and in animacy decision tasks that use the language-switching paradigm).

In contrast, given a control mechanism that acts on signals that emerge from the lexico-semantic system (external account), a language-switch cost relates to the competition between task schemata. Changing task schemata incurs a cost, thus a language-switch cost should appear when the task elicits separate, competing schemata. If the task can be accomplished with only a single schema, then one would not expect a language-switch cost, although language-signals might still affect responses 'paralexically'. As discussed earlier, it is assumed that bilinguals treat their two languages as distinct at the orthographic level of representation, but do not differentiate between languages at the semantic level (see also Thomas & Allport, 2000). Accordingly, it is supposed here that this disposition influences the types of task schemata set up by individuals when performing in an experiment. Specifically, when a bilingual participant is required to make lexical decisions, two task schemata are likely to be set up, since bilingual individuals presumably normally associate different computations with each of their two languages (though the possibility of a single task schema,

---

[37] Another means of presenting the other language indirectly, which was not used here, is to use letter strings that are nonwords in both the target and the non-target language, but which sound like words in the non-target language when pronounced according to target-language grapheme-phoneme rules (e.g., FUD pronounced according to German rules, sounds like the English word FOOD). Nas (1983) employed items of this type and showed that in an English-specific LDT, Dutch/English bilinguals were slower to reject letter strings that were nonwords in English but that sounded like Dutch words, compared to letter strings that were nonwords in English and had no status at all in Dutch.

given language-general instructions, shall not be ruled out entirely). In contrast, on the basis that a semantic task targets the language-neutral semantic representations (where bilinguals are unlikely to habitually invoke separate, language-specific computations during processing), it is likely that only a single task schema will be set up (unless task instructions made it necessary to distinguish between languages, e.g., a language-specific animacy decision task). Thus, in contrast to the internal account, the external account predicts a language-switch cost for a lexical decision task, but not for an animacy decision task (both using a language-switching paradigm).

On the basis of these considerations, three tasks were selected for the experiments presented in this thesis. The first task employed was a bilingual lexical decision task (Experiments 1a/1b and 2), which presented both languages to the participants by using the alternating runs paradigm[38]. The lexical decision task targets clearly defined mental representations that only require binary responses to single words. An additional advantage is that it is a widely used and well-established paradigm, making it easier to communicate findings and allowing comparisons to other research. The versions used here were language-general and language-specific LDTs. In the former, participants decide whether letter strings are words or nonwords, irrespective of language. In the language-specific LDT, participants decide whether a letter string is a word in a specific language, the language of decision being cued by factors such as screen colour or the position of the item on the screen (cf. Rogers & Monsell, 1995). The second task employed (Experiment 3) was an animacy decision task in which participants decided whether the word presented on the screen referred to an animate or to an inanimate entity. This task likewise permits both languages to be openly presented to the participant (i.e., presenting both German and English words as stimuli). An animacy decision task targets the language-neutral semantic representations, allowing clearly distinct predictions to be made with the internal and external accounts, given that the internal account proposes inhibition of non-target lexical representations (and the external account does not), which is critical, since activation of lexical representations is implicated in the processing for meaning. Furthermore, the animacy decision task is simple in that it requires binary responses to single words. The third task employed (Experiments 4a/4b), was an English-specific LDT,

---

[38] Rogers & Monsell (1995): two trials of language A followed by two trials of language B, etc. (e.g., English, English, German, German, etc.), yielding language-switch and language-non-switch trials in alternation.

meaning that the language of decision (English) was the same throughout - making this task essentially a unilingual LDT. Specific items can be introduced into such a task, that permit investigation of the influence of a language which is not only non-target, but not even openly presented in the entire experiment. That is, in contrast to the bilingual LDT, where both languages are openly presented and used, in the unilingual LDT only one of the two languages is relevant, and the second (non-target) language is presented only indirectly, in the form of interlingual homographs (such as MUTTER). In fact, as feedback from participants (Experiments 4a/4b) suggested, it appears that the non-target reading of such interlingual homographs is not necessarily consciously processed, yet nonetheless influences processing (see Chapter 5).

An additional advantage, common to all three paradigms, is that decision tasks afford the experimenter control over the choice and placement of stimuli. For example, in the bilingual LDT and in a bilingual ADT, stimuli can be arranged to have language switches on every other trial, allowing a within-block comparison between switch and non-switch trials (in contrast to the mixed block/pure block paradigm [used, for example, by Grainger & Beauvillain, 1987]). Finally, all three tasks provide two output measures (RT / accuracy), which can be analysed and used to interpret the data.

These single-word decision tasks are arguably unnatural, given that normal reading usually involves passages of text, and serves the function of conveying meaning (or indeed giving pleasure). However, the laboratory approach was justified by the consideration that simple issues need to be investigated in a controlled environment before it is possible to assess more complex matters (such as sentence comprehension), for which at present there would be too many variables. Certainly, the ultimate aim of these tasks is to illuminate processes taking place in natural reading, although the ecological validity of such tasks has, as yet, to be established. (For a detailed discussion of the ecological validity of laboratory tasks see Haberlandt, 1994.)

1.8 Summary and outlook

This chapter started off by showing why it is important to address the question how language is controlled in bilinguals. As a background, an overview of the unilingual language processing system was provided. On that basis, various models of bilingual language organisation, which dominated much of the early bilingual language research, were outlined.

It was shown that the early research provided important information regarding the 'architecture' of the bilingual lexico-semantic system. However, as was then discussed, these early studies were able to accommodate either language-switch costs or interlingual interference effects, but not both (and yet both are ubiquitous bilingual phenomena), because they focused solely on representation and left aside the question how the bilingual language processing system is controlled. Two models were then introduced that address precisely this shortcoming; these have here been referred to as the 'control models'. These control models provide two different accounts of how the bilingual language system may be regulated: within the lexico-semantic system, by inhibiting lexical representations (BIA model) vs. outside the lexico-semantic system, by use of task schemata (IC model).

The following chapters present four experiments designed to test different models of the bilingual language processing system, critically, the internal and external accounts. Two of the early accounts, which were influential in shaping the early research on bilingualism, are additionally assessed against the data collected here, to show how the accounts focusing on representation are unable to capture the range of bilingual behaviour that may be observed. One of these early accounts is the *input-switch* account (Macnamara & Kushnir, 1971), which was based on the notion of language-selective access to independent lexical systems. The other is the *associative connections* account (Dalrymple-Alford, 1985), which assumed connections between the two language systems, but specified the links between the two to be weak, making the language systems more independent than interdependent (though not going as far as the input-switch account, where the two are entirely separable).

The two 'control accounts' (the BIA and IC models), which are the main focus of this thesis, are based on the assumption of non-selective access to interdependent lexica. Both view phenomena like language-switch costs in terms of the way in which the lexico-semantic system is regulated, rather than as a consequence of the mental organisation of the two languages. However, the manner in which the bilingual language processing system is regulated, differs between these two accounts. Accordingly, each account provides a different explanation for the causes of phenomena such as language-switch costs and interlingual interference effects (such as interlingual homograph interference). As has been discussed in the course of this chapter, the language processing system cannot simply be regarded as a static storage system, from which information is retrieved when necessary. Humans react and adapt, depending on their circumstances. This will be shown clearly in Experiment 4a, which shows how interlingual interference can be modified over time, rather than being an unchangeable effect.

The necessary capabilities needed for a dynamic account of bilingual language processing are given by the BIA model (e.g., Dijkstra & van Heuven, 1998), which proposes a locus of control within the language processing system, and by the IC model (e.g., Green, 1998a, b), which (for language perception) proposes the primary locus of control to be outside the language processing system. In brief, on the *internal* account (BIA), regulation of the lexico-semantic system is posited to occur via the language nodes, which modulate the degree of activation of non-target language lexical representations. In contrast, on the *external* account (IC model), regulation is achieved by means of task schemata, that act differentially to the signals emerging from the lexico-semantic system. This, in fact, is an advantage which the external account has over the internal account, in that it models how signals from the lexico-semantic system may be linked to the units underlying responses. Otherwise, as was discussed above, in the sections pertaining to the BIA and IC models respectively, both models are able, in general terms, to provide a plausible account for the phenomena of language-switch costs *and* interlingual homograph interference.

This thesis presents different experiments exploring the phenomena of language-switch costs and interlingual interference, in order to test the two models (BIA and IC) that promise an account of not just one or the other phenomenon (as early accounts did), but that are able to accommodate both. Critically, the two control accounts represent an attempt to model how bilingual language processing may be controlled, which this introduction has shown must be possible. Previous models of bilingual language processing have not addressed this issue. The BIA and IC models are tested using different experiments, to ensure that evidence for or against either of them is not an artefact of the task used.

The first question to be addressed was whether task instructions could affect language-switch costs. Accounts focusing solely on representation make different predictions from accounts that assume some degree of control over the language processing system. Accordingly, Experiments 1a/1b and 2 used lexical decision tasks (alternating runs paradigm, Rogers & Monsell, 1995) to contrast the size of language-switch costs when under language-specific LDT instructions vs. language-switch costs when under language-general LDT instructions. Both of the control accounts predict larger language-switch costs given language-specific

LDT instructions. The input-switch and associative connections accounts do not (the details of this are given in the introduction to Chapter 2). In short, this first investigation could differentiate between the control accounts and the 'representational' accounts, but it could not very clearly make the critical distinction between the two control accounts (BIA / IC).

Experiment 2 then refined the lexical decision task, aiming to distinguish between the internal and external accounts, which Experiments 1a/1b had not been able to do clearly. To this end, not only task instructions, but additionally stimulus list composition (pure-HF vs. mixed frequency stimulus lists), and task order (language-specific LDT before language-general LDT and vice versa), were manipulated. The outcomes appeared to favour the external over the internal account.

Experiment 3 employed a different paradigm (animacy decisions) to provide further evidence that would allow one to distinguish more clearly between the internal and external accounts. This experiment again used words of both languages, presented with the alternating runs paradigm, providing language-switch and language-non-switch trials, which were crossed with response-type (response-switch / response-repetition). The internal account predicted a language-switch cost (given inhibition of lexical representations) regardless of response type, the external account predicted response type to modify language-switch costs. Significant language-switch costs were found for response-repetition trials, but not for response-switch trials, in line with the external, but not the internal account.

The final Experiment (4a/4b) contrasted the internal and external accounts from yet another angle: An English-specific LDT, that presented English words, nonwords and interlingual homographs (IHs). That is, a situation was created in which the second language was not only task irrelevant (as was also the case in the animacy decision task), but where the 'other' language was not even overtly presented. The critical data were IH interference effects (slower RTs on IHs than on matched controls), as well as 'carry-over effects' (slower RTs to pure English words placed directly after IHs than to pure English words placed directly after controls). Experiment 4a manipulated stimulus list composition, tracking IH interference and carry-over effects over time. Experiment 4b manipulated task instructions, again tracking IH and carry-over effects over time. The internal account predicted that IH and carry-over effects should be parallel; that is, if IH interference decreased, so should 'carry-over', since both IH interference and carry-over effects would reflect the degree of activation within the non-target lexicon. In contrast, the external account predicted that IH interference and carry-over would dissociate, since the former reflected task schema factors, while the latter reflected the degree of activation within the non-target lexicon. (See Chapter 5 for further details.) Experiments 4a and 4b both showed a clear dissociation between IH interference and carry-over effects, supporting the IC but not the BIA model.

At the beginning of this chapter, an example of interlingual interference in normal language use was given. The question was asked how it is possible that such interference may occur, or rather, how it is possible to (normally) *avoid* it. In the course of this chapter it has been shown that, to be able to answer this question, it is necessary to consider not just the architecture of the bilingual lexico-semantic system, but also how the system functions, specifically, how it is controlled. The two extant models which attempt to do this are the BIA model (e.g., Dijkstra & van Heuven, 1998) and the IC model (e.g., Green, 1998a, b). As the outline provided here has discussed, these two accounts are based on very different assumptions (control effected within vs. outside the lexico-semantic system). These assumptions are tested in this thesis.

II

Experiments 1a & 1b: Language-switch costs in language-specific and language-general lexical decision tasks

2.1 Introduction

The first two studies of this thesis (Experiments 1a and 1b) examine language-switching costs in bilingual lexical decision tasks 1) in the case where a word's language is relevant to the decision ('language-specific' lexical decision) and 2) in the case where a word's language is irrelevant to the decision ('language-general' lexical decision). Past studies showed that a cost is incurred for switching between languages (e.g., Kolers, 1966; Macnamara & Kushnir, 1971; Meyer & Ruddy, 1974; Dalrymple-Alford, 1985; Thomas & Allport, 1995). As discussed in Chapter 1, various accounts have been put forward as to where this cost for switching languages arises. Macnamara & Kushnir (1971) suggested that an *input-switch* has to be set, channelling the input into the correct lexicon. Every time there is a language-switch, the input-switch needs to be re-set, entailing a time cost (the language-switch cost). Macnamara & Kushnir (1971) supposed lexical access to be selective (only one lexicon accessed at a time), and the two language systems to be independent of one another (though cf. Chapter 1 for evidence to the contrary). Dalrymple-Alford (1985) suggested stronger *associative connections* between words of the same language, than between words of different languages. As discussed in Chapter 1, this proposal presumably employs a mechanism of spreading-activation. In a task requiring lexical identification, there is an advantage for words with stronger associations (compared to words with weaker associations), in that it will be faster to activate above threshold a lexical representation that has already received some activation (through spreading-activation), compared to a lexical representation which has not. On language-switch trials, reaction times (RTs) should be delayed compared to RTs on non-switch trials, because associative connections (between the previously seen word and the current word) are weaker in the former case than the latter, and so there is less 'pre-activation' of the target representation.

The *internal* and *external* accounts propose still different explanations of the language-switch cost, and also provide a means of exerting control (which the input-switch and associative connections accounts do not supply).

An *internal account* (Bilingual Interactive Activation [BIA] model) relates language-switch costs to the activation and suppression patterns within the lexico-semantic system ensuing upon bottom-up input (see Chapter 1). Lexical representations activate 'their' language-node, which in turn inhibits lexical representations of the other language (e.g., when processing an English word, the activated English language-node inhibits German lexical representations). The authors of this account assume that a previously active language-node does not deactivate itself instantaneously (cf. Dijkstra & van Heuven, 1998). For example, on the trial following an English item, the English node is still active (suppressing German lexical representations). If the next item is German, the German language-node needs to achieve greater activation than the English node. Changing the activation balance takes time. Thus RTs on language-switch trials should be slower than RTs on non-switch trials. (Note, Dijkstra and colleagues have not yet tested their model on language-switching tasks.)

Green's (1998a, b) Inhibitory Control model proposes an *external account.* On this account a language-switch cost arises at the output from the lexico-semantic system, where task schemata act differentially to signals emerging from the lexico-semantic system. A lexical decision task schema links signals indicating "a word" to the units underlying a "yes" response, and signals indicating "a nonword" to the units underlying a "no" response. For bilinguals, words are words of either one language or another, providing not just 'wordness' but also 'language' information (e.g., "it is a word; it is English" / "it is a word; it is German"). This is likely to be reflected in their processing of tasks. Furthermore, one may assume that the architecture of task schemata is influenced by the instructions given. Instructions such as "when the screen is blue decide if it is a word in English or not; when the screen is yellow decide if it is a word in German or not" encourage establishing two task schemata, one to link "German word" signals to 'yes' units, one to link "English word" signals to 'yes' units. Since only one schema can drive a response, a given schema must dominate. For a language-specific lexical decision task, task instructions require responding 'yes' solely to (for example) German words on German trials, and to successfully reject German words should they appear on trials designated as English ('wrong language words'). To achieve a given schema being dominant, the two schemata are proposed to be in strong mutual inhibition (i.e., the selected task schema strongly inhibits the unselected one). A change of target language requires a change of task schema. In contrast, when making a decision in the same language as on the preceding trial, the same task schema can be used again. Since the reconfiguration process takes time, RTs on language-switch trials are slower than those on non-switch trials.

Task set reconfiguration appears to be mainly stimulus driven (e.g., Allport, Styles &

Hsieh, 1994; Rogers & Monsell, 1995; Meiran, 1996; Meuter & Allport, 1999) and not so much under voluntary control. Thus even though participants may know what language is required on the next trial (as would be the case in a regular switching paradigm where language is made relevant), they apparently cannot fully reconfigure the system in anticipation (cf. Rogers & Monsell, 1995; Monsell, 1996). The task cue (e.g., screen colour) indicates which task schema is required, boosting its activation level. Additionally, signals from the lexico-semantic system (e.g., language signals) will presumably feed into the process, activating the schema they relate to (e.g., German signals activate a German task schema).

Given instructions to "accept if it is a word, either in German or in English" (a 'language-general task') individuals could establish a single task schema, since in such a task there is by definition no possibility of 'wrong language' words. However, it may well be that bilinguals are used to associating separate processes with each language at the level of word forms (e.g., separate syntax, separate orthographic constraints, separate phonologies). It may well be that it is difficult (or perhaps even impossible) to overcome such a practice effect for the purpose of a half-hour experiment (cf. also Thomas & Allport, 2000). Given language-general lexical decision instructions, the mutual inhibition between the two schemata should be weaker than given language-specific lexical decision instructions. Another possibility is for both schemata to be jointly active, the last-used being somewhat more active than the other, entailing some delay when switching languages. In either case, the switch-cost should be smaller in a language-general than in a language-specific LDT.

In the case that, after all, a single task schema were set up, would bilinguals avoid a language-switch cost? It seems very unlikely that bilinguals can ignore or suppress language signals (see also Kroll & Dijkstra, 2001). Such signals could feed into the decision process paralexically (see Chapter 1). That is, detecting a change of language could signal "a change has occurred" to the individual, indicating that the response which was previously *not* used is required on the current trial. If there was a language-switch but no response-switch (e.g., an English word followed by a German word, i.e., two consecutive 'yes' trials), then 'no' would have been activated erroneously on the basis of the 'language-change' signal. This would need to be revoked, 'yes' units would need to be activated sufficiently higher than erroneously activated 'no' units. In short, RT can be delayed (or errors made) on language-switch trials even when there is a single task schema. Note, such paralexically determined response-competition could certainly be concurrent with task schema competition, though it would not necessarily be possible to tell the two apart. To do so one would need to attempt to create a situation in which there is likely to be just a single task schema. Any remaining switch-cost would presumably then be attributable to paralexically modified response-competition. This is addressed in more detail in Chapter 4.

Experiments 1a and 1b are lexical decision tasks (LDTs) in which switching between the languages was regular and predictable. There were no interlingual homographs, so as to avoid confusing participants about the language of trials, which were to be experienced as they had been designated by the experimenter (switch/non-switch). The alternating runs paradigm (Rogers & Monsell, 1995) was used in both experiments. In this paradigm switch and non-switch trials alternate, since items are arranged such that there are two items of language A, then two of language B, then two of language A, etc. Thus language-switch and language-non-switch data are collected from the same experimental block. With views to levels of arousal and fatigue, this was considered preferable to a paradigm comparing mixed block data (language switches on every trial) with pure block data (all items are of one language only) - cf. Shafiullah & Monsell (1999). In this initial investigation, there was a balance between high and low frequency items, since the relative proportions of HF vs. LF items can affect lexical decision performance (for unilingual data see, for instance, Glanzer and Ehrenreich, 1979; Gordon, 1983; Dorfman & Glanzer, 1988). Manipulations of the proportions of HF and LF items will be addressed in Chapter 3.

In a 'language-specific' LDT there is a clear assignment of languages to trials (language cues are given) and participants are asked to make lexical decisions in the language of the current trial ("is this a word *in English*?" on 'English trials'; "is this a word *in German*?" on 'German trials'). In a 'language-general' LDT there is no assignment of languages to trials and participants need only make lexical decisions as a unilingual would - distinguishing simply between words and nonwords (i.e., "is it a word?"). The items were identical in Experiments 1a (language-specific LDT) and 1b (language-general LDT), and both used the alternating runs paradigm, allowing a direct comparison between the two tasks.

2.1.1 Predictions

First the predictions made by the *external* account are presented. Subsequently those made by the *internal*, the *input-switch* and the *associative connections* accounts are outlined.

(1) *Words.* a) It was predicted that the language-switch cost would be greater in the language-

specific compared to the language-general task. This is based on the idea that the language-specific task instructions clearly indicate the need to differentiate between languages (e.g., it would be wrong to respond 'yes' to a German word on an English trial). Thus there should be two separate task schemata with strong mutual inhibition. In the language-general task the mutual inhibition between schemata is likely to be weaker (or they are jointly active). Consequently, the switch-cost should be smaller. b) There should still, however, be a switch-cost in the language-general task, because there are still two schemata to switch between, though in this case that is easier. (If there were only a single schema, then there could be a switch-cost related to language signals [specifically: language-change signals], feeding into the decision process paralexically.)

(2) *Nonwords.* a) For the same reason as for words (two strongly mutually inhibited task schemata), there should be a clear switch-cost for nonwords in the language-specific task. In contrast, in the language-general task there may be no language-switch cost on nonword trials, for two reasons. 1) For nonwords language signals are likely to be weak, there being representations in the lexicon only of real-word neighbours, but none of the letter strings themselves. Therefore, any response-competition component of switch-costs (via paralexical decision processes) should be small. 2) A nonword could be adequately rejected by either task schema. Hence, it may be that on a language-switch trial a nonword is correctly rejected even when there has been no switch of task schema. b) If an effect of nonword type is found[39] (faster RTs for unique than for non-unique nonwords), then this should be comparable in the language-specific and the language-general tasks (whatever attribute of unique nonwords gives them a processing advantage over non-unique ones, should be equally available whether there are two task schemata with strong mutual inhibition, or two that are weakly inhibited / or jointly active, or even just one single schema).

---

[39] In a purely English LDT, Altenberg & Cairns (1983) found English/German bilinguals to be significantly slower (by an average 51ms) to reject 'non-unique' nonwords than 'unique' (English) ones. They interpreted this as reflecting influences from the non-target language (German) on processing in the target language (English). However, English *unilinguals* were also significantly slower (by an average 61ms) on non-unique compared to unique nonwords, and bilinguals, carrying out the task in their weaker L2 (German), showed *no* nonword type effect. This suggests that some orthographic response strategy allows faster responses on unique compared to non-unique nonwords, but that it is not (or at least not only) a question of cross-language influences. In fact, if there were such influences, then one should expect the nonword type effect to be larger for bilinguals than for unilinguals. That was not the case. However, this new interpretation should not be seen as support for the input-switch's 'separate lexica' viewpoint. After all, nonwords are not represented in a person's lexicon, and so, not finding cross-language effects with nonword stimuli does not rule out cross-language influences for other stimulus types.

Note, since it is uncertain what the basis of the nonword type effect is (see footnote 39), the main focus of Experiments 1a and 1b (whose aim is to shed light on the locus of control) is on the word data. The nonwords are necessary as a contrast to words in the lexical decision task. Manipulating nonword type was important to hinder participants using an orthographic strategy for their lexical decisions to words (see Thomas & Allport, 2000, with respect to Grainger & Beauvillain, 1987). Whether nonword type (as used here) can also give information about language processing (i.e., the question of selective and non-selective access) is not clear. Thus the nonword data here serve mainly as a qualifying backup to the word data, they do not pursue an agenda of their own.

The *internal account* (Dijkstra & van Heuven, 1998) predicts a) a larger language-switch cost in the language-specific compared to the language-general task (as did the *external* account). It does so by proposing that the top-down suppression of non-target lexical representations is greater in the language-specific than in the language-general task, because in the former the instructions require processing in only one language at a time. It thus predicts an interaction between task and trial type for both words and nonwords. However, it predicts that there should still be a switch-cost for nonwords in the language-general task (contrary to the *external* account), because the language-node activation balance needs to be changed on a language-switch trial and, despite paralexical processes, a response cannot be given until this is achieved. b) If the nonword type effect entails a component of cross-language influences, then the internal account would predict a smaller nonword type effect in the language-specific task than in the language-general task (because the non-target language is more strongly suppressed in the former task, with correspondingly less chance of interference from non-unique nonwords' real-word neighbours in the non-target language). If nonword type is (mainly) something other than cross-language effects, then there should be no interaction between nonword type and task.

The *input-switch* account makes the following predictions.

Words. (Note that real words used here were orthographically legal in either language, i.e., they themselves provided no cues to use for re-setting the input-switch on a language-switch trial.) Language-switch trials involve a re-setting of the input-switch, non-switch trials do not. Thus this account predicts slower RTs on language-switch than non-switch trials. The size of the switch-cost should be modified by the task. In the *language-specific task*, colour can cue the resetting of the input-switch on language-switch trials. Having detected the colour change, the input-switch is re-set, the corresponding lexicon is searched, and a response can be given. In the *language-general task* there is no colour cue, and the words provide no clear orthographic cues which might help to determine which lexicon is the relevant one. Assuming that the input-switch remains in the previously used position until there is evidence that it needs to be re-set, a language-switch trial entails that the input is first 'fed into' the lexicon in use until now, and, upon failure to find a match there, 'fed into' the other lexicon. In sum, for the orthographically neutral words used here, the input-switch account predicts that language-switch trials have longer delays in the language-general than in the language-specific task, since in the language-general task a language-switch trial not only involves a re-setting of the input-switch, it additionally entails searching two lexica rather than one (on the assumption that the input-switch remains in the position last used). Thus, contrary to the external account, the input-switch hypothesis predicts a larger language-switch cost in the language-general than in the language-specific task.

Nonwords. The main predictions of the input-switch account with regard to nonword data can be summarised as follows. *Language-specific task*: In this task there are colour cues which can be used to help set the input-switch. The resulting predictions are: a) There should be a switch-cost (cued by the screen colour, the input-switch is re-set on switch trials, but not on non-switch trials). b) Overall faster RTs for unique compared to non-unique nonwords (because the unique orthography of the former can aid the resetting of the input-switch on language-switch trials). I.e., a nonword type effect is related to the switching mechanism, it does not indicate cross-language influences. c) An interaction between language-switching and nonword type (because while there may be similar RTs for the two nonword types on non-switch trials, the orthography of the unique nonwords can aid the re-setting of the input-switch on language-switch trials, which the non-unique nonwords cannot achieve). *Language-general task:* In this case there are no colour cues which can be used to set the input-switch on language-switch trials. The only cues available are orthographic. Only the unique nonwords have such cues (e.g., PF in PFITZE can set the input-switch towards German; since the letter combination PF violates English orthographic rules, there is no need to search for this item in the English lexicon). Non-unique nonwords could potentially be words in either language and so both lexica must be searched, regardless of the trial's designation[40]. The

---

[40] There being no colour cues, and 1/4 of all items being orthographically neutral nonwords, participants are likely to be unaware of the trial designations (switch / non-switch).

predictions are: a) A small switch-cost, if any at all (non-switch trials are faster than switch trials only for unique nonwords). b) Faster RTs for unique than for non-unique nonwords (the latter need to be searched for in both lexica, the former do not). c) An interaction between trial type and nonword type, with slower RTs on language-switch than non-switch trials for unique nonwords, but no difference between switch and non-switch trials for non-unique nonwords (they are effectively always switch trials, whether designated so or not).
Taken together, there should, for nonwords, be a smaller switch-cost in the language-general compared to the language-specific task (as predicted by the external account). Contrary to the external account, there should be an interaction between task and nonword type, due to a greater difference between unique and non-unique nonwords in the language-general than in the language-specific task (because of the non-unique nonwords which are particularly slow in the language-general task, having no orthographic cues to direct the input-switch to a given lexicon).

Dalrymple-Alford's (1985) *associative connections* account allows predictions about words. There being no representations for nonwords in peoples' lexica, it cannot easily make predictions for nonwords. In any event, since word data are most relevant for the current purposes, these shall be focused on.

The colour cue is assumed to activate the required language's representations, and presumably to begin having this effect while the letter string is still being analysed. Thus, in the language-specific task, all representations of the corresponding language, including that of the presented letter string, receive some activation early on. In contrast, in the language-general task, the representation of the presented letter string only receives activation once the visual features have been analysed. Such preferential activation (language-specific task) would entail that it takes less time for a representation to reach threshold, than in the case where there are no colour cues (language-general task). The consequence for language-switch trials is an advantage for language-switch trials in the language-specific task compared to those in the language-general task. In sum, this account predicts (contrary to the *external* account) that there should be a smaller language-switch cost in the language-specific compared to the language-general task.

2.1.2 Signals available in lexical decision tasks and their usage

Apart from 'language' signals, a presented word gives rise to 'wordness' signals (cf. Monsell, Doyle & Haggard, 1989; Grainger & Jacobs, 1996). Both can in principle be used to make lexical decisions. However, in a language-specific task the participant is meant to respond 'yes' only if an item is a word in the required language, not simply on the basis that it is a word. One way of enforcing compliance with language-specific instructions would be to place some language A words on trials designated for language B ('wrong language' items). This would provide a behavioural check (Thomas & Allport, 2000), in that individuals would need to reject such items. However, the aim here was to compare a language-general with a language-specific task. By definition, in a language-general task a 'wrong language word' cannot exist, there being no language specification of trials. To include such items in one task but none in the other would have made the two tasks less comparable. For an initial investigation into the hypothesis that task schemata, and not inhibitory processes within the lexico-semantic system, lie at the bottom of the language-switch cost, it was important to keep the tasks themselves as 'pure' and similar as possible.

Without 'wrong language' items, there is nothing to hinder participants from responding 'yes' as soon as a 'wordness' signal is available (i.e., 'reconstruing' the language-specific task as a language-general one, as de Groot, Delmaar & Lupker [2000] call it). If participants did this, then the language-switch costs in the two tasks would not differ. But since participants were new to lexical decision tasks, it was thought that they would comply with the task instructions.

Experiment 1a: language-specific LDT

2.2. Method

2.2.1. Participants

20 German/English bilinguals (7 males, 13 females), drawn from the student population of University College London, participated. None had previous experience with lexical decision tasks. Their mean age was 27.3 years (4.0), range 21-33. All participants had normal or corrected vision; colour vision was also normal for all. Participation was voluntary and remunerated at the standard departmental rate.

Language proficiency:

All participants grew up with German and started to learn English at school as their second language at a mean age of 10.6 years (2.0). At the time of the experiment the participants had been using English continuously between 1 and 14 years (mean = 4.0, *SD* = 3.5).

In order to gain a measure of their language proficiency, participants carried out, following the experiment, two computerised word recognition tests ('LLEX' - Meara, 1994) and completed a language background questionnaire – see Appendix 1. In the LLEX test (which has separate tests for several languages, among them English and German), individuals were presented with real words and nonwords (one by one) and asked to indicate whether they knew the item as a word of the current test language or not. LLEX results are based on accuracy only, and so individuals were told that they no longer needed to make speeded responses (as in the LDT), but encouraged to not spend too much time on each individual item either. Results (out of a maximum of 100) indicate the level of proficiency. For English the results of the LLEX test ranged 76-98 out of 100 (mean = 87.4, *SD* = 6.2). For German the range was 72-99 (mean = 89.7, *SD* = 6.8), suggesting that these bilinguals were reasonably balanced. However, the results of the self-rating (within the language background questionnaire) showed a somewhat different picture. Here participants were asked to assess their relative proficiency in German as compared to English (out of a combined total of 100 points for both languages together). The results were somewhat higher estimates for German than for English (German: mean = 57.0, *SD* = 9.4). Participants reported their current use of the two languages (also out of a combined total of 100 points) to be predominantly English: the average score for English was 66.8 (*SD* = 26.3). Participants were given a choice in which language to fill in the questionnaire - 12 chose the German, 8 the English version.

2.2.2 Design

There were two within-subjects designs, one for words and one for nonwords. The word analysis had the within-subjects factors Language (English / German), Trial Type (switch / non-switch) and Frequency (high frequency / low frequency). For nonwords the design was identical except that the factor Frequency was replaced by Nonword Type (non-unique/unique). Individuals were shown an equal number of word and nonword stimuli.

Trial Sequences:

The alternating runs paradigm was used, so that language switches occurred on alternate trials (EEGGEE etc.)[41]. Two colours were used to cue the respective languages (counterbalanced across participants). A change in screen colour signalled a language-switch. Each type of

---

[41] E = English; G = German.

stimulus (HF/LF words in German and in English, non-unique/unique nonwords for German and for English trials) occurred once every 8 trials. Eight sets of such eight-trial sequences (with a different order of stimuli in each case) were combined to form an experimental block of trials (i.e., 64 experimental items per block). Five such experimental blocks, permuted in different orders, combined to yield the set of experimental trials (i.e., 320 items). All items (HF/LF words, unique/non-unique nonwords) were balanced over language-switch and non-switch trials. Items seen on switch trials by one half of the participants were seen on non-switch trials by the other half. Each experimental block was preceded by a single filler trial to ensure a clear designation for the first measured trial in each experimental block (i.e., 65 items per block and 325 items in total in the experiment). These filler trials were excluded from the analysis. Words were presented in one language only to a participant (no translations) to avoid repetition-priming effects, and successive items were (by judgement of the experimenter) semantically and associatively unrelated to avoid associative-priming effects (see e.g., de Groot & Nas, 1991, for cross-language repetition- and associative-priming effects). Half the participants saw one set of stimuli and half the other matched set of stimuli (further details in *Materials* below).

2.2.3 Materials

a) Words:

Two sets of words (set A, set B) comprised a total of 160 words each (the filler trials presented at the start of each block were additional to these items). In each set, 80 words were English and 80 words were German. In each case, half the words were of high word frequency (HF) and half were of low word frequency (LF) as determined by the Kučera & Francis (1967) norms for English that comprises a corpus of one million words and the Ruoff (1990) norms for German, based on a corpus of half a million words (at this point there was not yet access to the CELEX database - Baayen, Piepenbrock & van Rijn, 1993). The words in each set were also matched for syllable length and letter length across the two languages (see Table 2.1).

Words were chosen to be orthographically legal in either language, as far as constraints owing to frequency bands allowed. No words were orthographically illegal, but some were perhaps less usual for the other language (e.g., LEUTE, BRAIN). Constraints on the orthographic properties of words were imposed to preclude individuals from using an orthographic strategy for their decisions, such as responding 'yes' upon recognising a letter combination unique to one language without having fully identified the item (see Thomas & Allport, 2000, on Grainger & Beauvillain, 1987). Further to discourage such a strategy, orthographic properties of nonwords were manipulated (see below).

Cognate homographs (sharing spelling and meaning across languages, e.g., German/English: FINGER) and non-cognate homographs (sharing spelling but not meaning across languages, e.g., German/English: TAG ['day' in German]) were excluded for several reasons. 1) To avoid confusing participants as to the language of the current trial in the language-specific task. 2) On such items it would not have been clear which reading (English or German) had served as the basis for the 'yes' response (assuming a correct reply). This would further have meant that there would have been less control over which trials were 'switch' and which were 'non-switch'. 3) Such items could have been perceived as 'wrong language' items (Thomas & Allport, 2000), which were deliberately excluded in this design.

Words had to conform as far as possible to the variety of matching parameters outlined above. Regarding the English words, it was additionally important that they would be known as English words by the bilingual participants for whom English was their second language. Errors due to processing slips are analytically interesting, but errors due to lack of knowledge are not experimentally informative. For these reasons some items included in the stimulus lists, that complied with the orthographic, length and 'knowledge' requirements, had frequencies clearly higher than the average. Such items cause the mean frequency values of German and English words in Table 2.1 to look somewhat unmatched, and are reflected in the varying *SD*s. But note also, that any frequency counts are in any event only an estimate of what the subjective frequency might be for each participant, whereas properties such as word length are objective, so that matching for those parameters should be well balanced.

Table 2.1

Mean frequencies (English: Kučera & Francis; German: Ruoff), mean letter length and mean syllable length (SD in brackets in each case) for the word stimuli, as a function of language and word frequency band (HF / LF), for set A and set B separately. CELEX values (Baayen, Piepenbrock & van Rijn, 1993) are included for completeness' sake only[1]; they were not available at the time of stimulus assembly.

		Set A			
		.Frequency		Length	
Language	**Word Type**	**Kučera & Francis**	**CELEX**	**Letter**	**Syllable**
English	**HF[2]**	135.2 (165.4)	118.1 (172.7)	4.50 (0.8)	1.50 (0.5)
	LF[3]	3.0 (1.7)	3.6 (2.7)	4.47 (0.7)	1.50 (0.5)
		Ruoff			
German	**HF**	97.1 (118.4)	123.6 (171.8)	4.48 (0.9)	1.50 (0.5)
	LF	1.0 (0.0)	4.3 (5.0)	4.52 (0.6)	1.50 (0.5)
		Set B			
		Frequency		Length	
Language	**Word type**	**Kučera & Francis**	**CELEX**	**Letter**	**Syllable**
English	**HF**	159.0 (257.3)	135.2 (283.0)	4.52 (0.8)	1.50 (0.5)
	LF	3.0 (1.5)	4.3 (5.6)	4.48 (0.7)	1.50 (0.5)
		Ruoff			
German	**HF**	92.4 (133.1)	78.9 (127.9)	4.50 (0.8)	1.50 (0.5)
	LF	1.0 (0.0)	19.2 (47.5)	4.53 (0.7)	1.50 (0.5)

[1] CELEX values could be regarded as closer to the present participants' subjective ferquencies than K & F, being compiled in the 1980ies rather than the 1960ies, having a larger corpus (18 vs. 1 million), and being based on British English rather than American English materials.

[2] HF = high frequency; [3] LF = low frequency

b) Nonwords:

Nonwords were constructed from real words (not otherwise used in the experiment) by changing one vowel. The words from which the nonwords were derived were similar to the word stimuli in that half of the 'base words' were HF and half LF (for 'unique' and 'non-unique' within each language respectively). Including both unique and non-unique nonwords precluded participants from using an orthographic strategy for replying to words (see Thomas & Allport, 2000, with respect to Grainger & Beauvillain, 1987). The unique nonwords were legal only in one language on the basis that they contained a consonant cluster illegal in the non-target language (e.g., PFLOG for the German stimuli; SWERM for the English stimuli). The non-unique nonwords were derived from German/English interlingual homographs, and were legal in both the target and the non-target language (e.g., HOND). All nonwords were balanced (among each other) for the position of the changed vowel in the word (letter position). The unique NWs were further balanced among each other for the position of the consonant cluster illegal in the non-target language (word-initial, word-medial or word-final - as the clusters vary in terms of the numbers of letters, actual letter positions cannot be given, hence the more loose terms 'initial', 'medial' and 'final') - see Table 2.2.

Nonword-Type rating study:

The contrast between the two types of nonwords was assessed by having 12 English unilinguals and 12 German/English bilinguals (not participating in the computerised part of the study) rate items for their 'word-likeness' in English and German. Each rater was given two lists of nonwords. A "German" list (comprising 'German' unique and non-unique nonwords), and an "English" list (comprising 'English' unique and non-unique nonwords). The English unilinguals rated their lists (English and German) for wordness in English. The German bilinguals rated their lists (English and German) for wordness in German. Ratings were given on a scale of 1-8. For half of the raters in each language group, "1" equalled "could be a word in ...."[42] and "8" equalled "couldn't possibly be a word in ..."[42]. For the other half this was reversed. In the results given below, the figure "1" indicates a low wordness rating, "8" a high wordness rating.

Results showed items to be distinct from one another in the direction suggested by Altenberg & Cairns' (1983) findings. **English unilinguals** rating the "German" list (for English word-likeness) rated non-unique nonwords as more word-like than unique (German) nonwords (the medians were 5 and 2, respectively). That is, violations to English orthography (items orthographically unique to German) induced low wordness ratings for unilingual English speakers. In contrast, the "English" list (rated for English word-likeness) elicited comparable ratings for non-unique and unique (English) items (median value 4 in each case). **German bilinguals** rating the "German" list (for German word-likeness) gave high ratings to unique (German) nonwords (median = 7) and slightly lower ratings for non-unique nonwords (median = 5), suggesting that they were sensitive to the discriminating orthographic features. Rating the "English" list (for German word-likeness), they rated non-unique nonwords as more word-like than the unique (English) nonwords: the medians were 6 and 2, respectively.

Two matching sets of 160 nonwords were created from this pool of nonwords. In each set there were 80 nonwords for the English trials (half were non-unique, half were unique) and 80 nonwords for the German trials (half were non-unique, half were unique). Table 2.2 provides the descriptive details.

---

[42] "... English" for the English ratings, "... German" for the German ratings.

Table 2.2

Mean letter length, mean syllable length, mean position of vowel exchange (given as letter position) and number of items per illegal cluster position (given as general position in the word: word initial, medial, final) for the nonword stimuli, as a function of language (English / German) and nonword type ('non-unique'[1] / 'unique'[2]), for Set A and Set B separately. Standard deviations are given in brackets.*

Set A

		Length		Nonword Determinants	
Language	**Nonword Type**	**Letters**	**Syllables**	**Position of Vowel Change**	**Position of Illegal Cluster***
English	**Non-Unique**	4.50	1.50	2.28	_
		(0.7)	(0.6)	(0.6)	
	Unique	5.00	1.50	2.28	initial: 16 items
		(0.7)	(0.5)	(0.8)	medial: 12 items
					final: 12 items
German	**Non-Unique**	4.48	1.50	2.30	_
		(0.7)	(0.6)	(0.7)	
	Unique	5.00	1.50	2.30	initial: 13 items
		(0.4)	(0.5)	(1.0)	medial: 14 items
					final: 13 items

Set B

		Length		Nonword Determinants	
Language	**Nonword Type**	**Letters**	**Syllables**	**Position of Vowel Change**	**Position of Illegal Cluster***
English	**Non-Unique**	4.50	1.50	2.28	_
		(0.7)	(0.5)	(0.7)	
	Unique	5.00	1.50	2.30	initial: 16 items
		(0.5)	(0.5)	(0.6)	medial: 12 items
					final: 12 items
German	**Non-Unique**	4.48	1.53	2.30	_
		(0.7)	(0.6)	(0.7)	
	Unique	5.00	1.50	2.30	initial: 13 items
		(0.4)	(0.5)	(1.1)	medial: 14 items
					final: 13 items

* Since 'illegal clusters' differ in length (2-4 letters), means for letter positions of illegal clusters cannot sensibly be reported, hence the total number of nonwords falling into each category (initial/medial/final) are reported.

[1] 'non-unique': orthographically legal in either language

[2] 'unique': orthographically legal only in the target language

2.2.4 Apparatus & Procedure

The experiment was run on an IBM-compatible PC in a small cubicle. All participants were tested individually. Stimuli were presented in white upper-case letters, centred on the screen which was either blue or yellow. All words were capitalised with a view to preventing language cues in the language-general task (in German nouns are capitalised). A two-button response box, marked "+" (word) and "-" (nonword), registered responses. Buttons were simply marked + and - in order to be language-neutral. Individuals pressed the space bar of an adjacent computer keyboard to initiate a block of experimental trials.

Participants saw written instructions (in English) on the computer screen, and also received verbal instructions (in German) from the experimenter. Instructions were in both English and German to ensure that both languages were active. Participants were informed that when a letter string appeared on the screen they were to decide as quickly and as accurately as possible whether or not it was a word in a specific language. Half the participants were instructed to determine whether or not the letter string was a German word when the screen colour was blue, and to decide whether or not the letter string was an English word when the screen colour was yellow. The remaining participants had the reverse assignment of colour and language. Using only their dominant hand, all participants pressed the button marked "+" with the forefinger, and the adjacent "-" button with the middle finger. (This experiment does not compare speed of responding to words vs. nonwords, and word data are more important for the issue at hand. Hence it was deemed permissible to not counterbalance the assignment of responses to fingers.) The other hand was used to press the space bar on the computer keyboard (to move through the instructions on the screen and to initiate a block of trials).

After initiating the block of trials, a letter string was presented on the coloured computer screen and remained there until the individual responded by pressing one of the two response keys. 'Reaction time' (RT) was the interval between stimulus onset and the button-press. One second after the button-press, the next stimulus appeared on the screen. Such a long response-stimulus interval was chosen so that any cost found for switching between languages could be attributed to effects of the stimuli, rather than to incomplete preparation (see Allport, Styles & Hsieh, 1994; Rogers & Monsell, 1995; Meiran, 1996; Allport & Wylie, 2000; Meiran, Chorev & Sapir, 2000). The language for lexical decision changed on alternate trials, indicated by a change of the screen colour. A short bleep signalled an error.

Each participant completed a practice block of 16 trials, followed by 325 experimental trials (including 5 filler trials, one at the beginning of each block of 64 trials). Individuals were encouraged to pause between experimental blocks. After finishing the experiment, individuals carried out the LLEX tests and filled in the language background questionnaire.

2.3 Results & Discussion of Experiment 1a

2.3.1 Data treatment

Responses on which errors were made, as well as particularly fast or slow responses ('outliers', +/- 2 *SD* from each participant's cell mean), were excluded from the RT analyses[43]. For words the error rate was 9.4%, for nonwords it was 9.3% (9.4% overall). The outlier rate was 3.6% for words and 3.9% for nonwords (3.7% overall). Analyses were carried out both on means and medians of correct RTs. In the following the means analyses are reported. The pattern of the median analyses followed that of the means analyses, for which reason they are not reported separately[44]. Where degrees of freedom are lowered and variable in the item analyses, this is because removal of errors and outliers reduced some cells to zero. Analyses were also performed on (arcsine transformed) error proportions. Analyses were performed by subjects (F_1) and by items (F_2). Coleman (1979) recommends reporting three F ratios (see also Forster & Dickinson, 1976; Raaijmakers, Schrijnemakers & Gremmen, 1999). The F_1, which treats participants as the generalisation variable; the F_2, which treats items as the generalisation variable, and the min F', which treats both as simultaneous generalisation variables. The min F' statistics were calculated following Clark (1973, page 347): "If F_1 has n and n_1 degrees of freedom, and F_2 has n and n_2, then $i = n$, and j is the nearest integer...", using the following two formulae for the calculation

$$j = (F_1 + F_2)^2 \ / \ (F_1^2/n_2 + F_2^2/n_1)$$

$$\min F'(i,j) = F_1 \cdot F_2 \ / \ F_1 + F_2$$

Note, as Raaijmakers et al. (1999) state, item variability need not be controlled statistically if it has been controlled experimentally. This can be achieved by matching stimuli with respect to variables correlating highly with the dependent variable (e.g., matching for word frequency and word length when investigating RT). When this is done, they observe, an analysis by subjects would suffice. Despite this, F_2 and min F' analyses were carried out in this thesis as well. This was done against the possibility that items might not be matched by participants' subjective word frequencies, even when by existing frequency counts (e.g., Kučera & Francis, 1967) they were well matched. (This issue arises in Experiments 4a/4b.)

Inspection of RT and error data revealed no speed/accuracy trade-off. Mean values presented in the text and tables are all mean of means of the correct RTs (by subjects). The temporal

---

[43] Outliers were removed only for the means, but not the medians analyses.

[44] This similarity is mentioned because it is important. It indicates that the observed patterns (critically, language-switch costs), do not arise from slow responses on just a few language-switch trials.

switch-cost was calculated by subtracting the RT of non-switch trials from that of switch trials (the corresponding method was employed to establish the cost to accuracy).

2.3.2 Reaction Time (RT) and error analyses for word trials

For the analyses by subjects a fully within-subjects repeated-measures ANOVA with the factors Trial Type (language-switch / language-non-switch), Language (English / German) and Frequency (HF / LF) was carried out. For the analyses by items the factors Language and Frequency were between-subjects, all others were within-subjects. Table 2.3 displays mean correct RTs and error rates as a function of language, word frequency and trial type.

Table 2.3

Experiment 1a (language-specific LDT). Mean correct RT (ms), standard deviations (in brackets) and error percentage per cell for all words, as a function of language (English/German), frequency (high/low) and trial type (switch/non-switch). The switch-cost (ms) is included to facilitate reading the table.

		Trial Type						
		Switch			Non-Switch			Switch
Language	**Word Type**	**Mean RT**	**(*SD*)**	**% Error**	**Mean RT**	**(*SD*)**	**% Error**	**Cost (ms)**
English	**HF**[1]	783	(183)	1.0	669	(151)	1.3	114
	LF[2]	991	(242)	21.5	835	(192)	22.8	156
German	**HF**	753	(172)	2.8	663	(134)	2.5	130
	LF	856	(227)	10.5	747	(171)	13.3	109

[1] HF = high frequency [2] LF = low frequency

Most importantly, Table 2.3 shows consistently slower RTs on switch compared to non-switch trials, which is comparable for both languages (see also Figure 2.1). The following provides the analytical details of these and other effects. (Less important information is only briefly dealt with here. The corresponding analytical details are then presented in Appendix 3).

All three main effects were significant in the RT analysis, but not all in the error analysis (giving the impression that individuals placed themselves towards the accuracy end of the speed/accuracy continuum).

The main finding is that (consistent with all four accounts: *external, internal, input-switch* and *associative connections*) there was an effect of trial type: on average, individuals were 117ms slower on switch trials compared to non-switch trials [$F_1(1,19) = 51.56$, MSE =

10767.36, $p < 0.001$; $F_2(1,296) = 59.69$, MSE = 46505.64, $p < 0.001$; min F'(1,63) = 27.66, $p < 0.001$]. That is, a language-switch cost was found for German/English bilinguals, extending Thomas & Allport's (1995) earlier findings regarding English/French bilinguals. There were somewhat fewer errors on language-switch trials (9.0%) than language non-switch trials (10.0%), but this difference was significant only by subjects [$F_1(1,19) = 4.39$, MSE = 0.01, $p < 0.05$; F_2 and min F' < 1].

Consistent with their self-ratings of language proficiency, individuals were overall 63ms faster responding to German words than to English words [by subjects and items $p < 0.001$; on min F' $p = 0.01$]. They made fewer errors on German words (7.3%) than on English words (11.7%), but this was significant only by items [$p < 0.001$; F_1 and min F' < 1]. (See Appendix 3.)

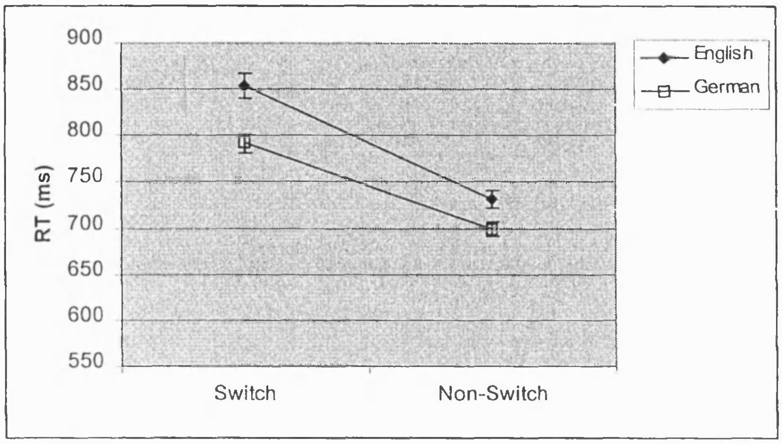
There was a main effect of frequency [all $p < 0.001$], such that RTs to HF words were faster and more accurate than those to LF words (140ms faster, 15.1% fewer errors). There was an interaction between frequency and language [RT: min F' $p = 0.062$, all others (RT and error) at least $p < 0.05$], with little difference (in terms of both RT and errors) on HF words, but a substantial difference on LF words (see Table 2.3 for RT and error rates, and Appendix 3 for statistical data). These data suggest a better knowledge of German (particularly regarding LF words), in line with the language background information presented in the *Method* section.

No other interactions were significant.

Figure 2.1 shows that, although RTs to English words were slower than RTs to German words, the effect of language-switching was the same for both; there was no statistically reliable interaction between language and trial type (i.e., there is no reliable evidence for asymmetric switch-costs). Since frequency did not modify the effect of switching, the values in Figure 2.1 are summed over high and low frequency words.

Figure 2.1

Mean correct RT (ms) of words (averaged over HF and LF) on switch and non-switch trials as a function of language (German [L1] / English [L2]) in the language-specific LDT (Experiment 1a). Error bars are standard error of the mean.



2.3.3 RT analysis for a subset of high frequency word trials

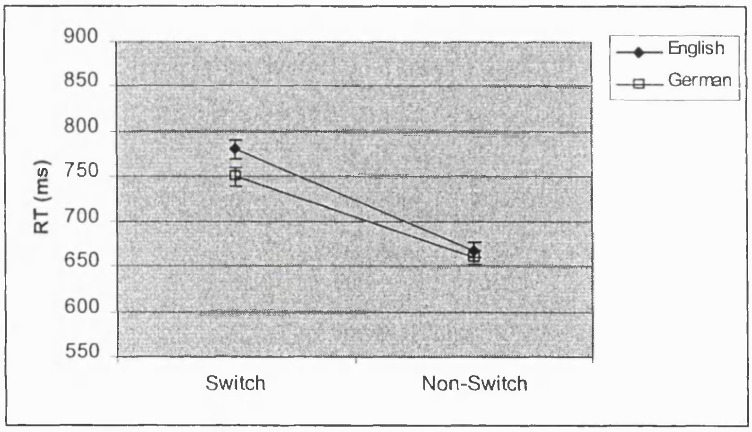
The effects of switching and language were also analysed for a subset of high frequency words, each of which was preceded by an HF word (of the same or the other language). This investigation had not been envisaged at the outset, so that the items in this set were not necessarily the same for all participants and given items were not balanced over switch and non-switch trials. For this reason, only the subject analyses could be carried out. The interest in this subset is that one can in this case be reasonably sure that the language-switch and language-non-switch trials are experienced by the participant as designated by the experimenter (where the prior trial involves a non-unique nonword, one cannot be entirely sure which task schema was employed when rejecting that non-unique nonword). Also, for HF words error rates are clearly lower than for LF words, hence the selection of HF words following other HF words, to ensure that the items of interest were preceded by correct responses and provided many correct RTs themselves. In fact, error rates for these critical trials were very small (3.5%), so that error analyses could not sensibly be carried out.

Since this additional investigation was not envisaged at the outset, there was only a small set of 'HF following HF' instances. There was one of each type per experimental block (i.e., one English HF word on a switch trial and one on a non-switch trial, the same for German), making a total of five instances of each type in the entire experiment per participant. These numbers are very small, and so the following results can only be seen as tentative indications.

As one might expect on the basis of the overall analysis (little difference between English and German on HF trials), there was no significant effect of language [$F_1 < 1$], but there was a substantial effect of trial type [$F_1(1,19) = 81.85$ MSE = 8920.74, $p < 0.001$]. The average switch-cost was 192ms (average RT on switch trials: 811ms; on non-switch trials: 619ms). There was no reliable interaction between trial type and language [$F_1(1,19) = 2.02$, $p = 0.17$] – i.e., the switch-cost appears to have been symmetric for this subset, just as it had been for the full word set. See Figure 2.2.

Figure 2.2

Mean correct RT of the HF subset on switch and non-switch trials as a function of language (German [L1] and English [L2]) in the language-specific LDT (Experiment 1a). Error bars are standard error of the mean.



2.3.4 RT and error analyses for nonword trials

For the analyses by subjects, the repeated-measures ANOVA comprised the within-subjects factors Trial Type (language-switch / non-switch), Language (English / German) and Nonword Type (non-unique / unique). For the analyses by items, Language and Nonword Type were between-subjects factors. Table 2.4 displays the relevant data.

Table 2.4

Experiment 1a (language-specific lexical decisions). Mean correct RT (ms), standard deviations (in brackets) and error percentage per cell for nonwords, as a function of language (English / German), nonword type (non-unique[1] / unique[2]) and trial type (switch / non-switch). Switch-costs have been included to facilitate reading the table.

		Trial Type						
		Switch			Non-Switch			Switch
Language	**Nonword Type**	**Mean RT**	**(*SD*)**	**% Error**	**Mean RT**	**(*SD*)**	**% Error**	**Cost (ms)**
English	**Non-Unique**	1020	(267)	9.0	891	(230)	12.0	129
	Unique	984	(274)	12.3	894	(247)	19.0	90
German	**Non-Unique**	1041	(271)	5.8	902	(192)	6.5	139
	Unique	908	(234)	4.5	874	(181)	5.8	34

[1] 'non-unique': orthographically legal in either language; [2] 'unique': orthographically legal only in the target language

The data pattern in Table 2.4 shows consistently slower RTs on language-switch than non-switch trials. Generally, RTs for unique nonwords are faster than those for non-unique nonwords (except for the English non-switch trials), but this is clearly evident only on language-switch trials. The following section provides the analytical details.

As Table 2.4 indicated, individuals were slower to respond (by an average 98ms) on switch than on non-switch trials, and made somewhat fewer errors (2.9% less) on switch than non-switch trials. However, neither difference was significant [Fs < 1 or $p > 0.05$]. The external and internal accounts had both predicted a switch-cost for nonword trials. It is possible that the data were too noisy for the trial type effect to be established statistically. (Indeed, Experiment 2 [Chapter 3] shows a significant language-switch cost on nonword trials in a language-specific task.)

RTs to non-unique nonwords were slower (by 49ms) than those to unique nonwords (average RTs were 964ms and 915ms respectively), but this was significant only by subjects [nonword type: $F_1(1,19) = 18.18$, MSE = 3245.31, $p < 0.01$; F_2 and min F' < 1]. (Removal of errors and outliers reduced some cells to zero in the item analysis. It is possible that with lower power this analysis did not reach significance above the data's noise.) Error rates were slightly higher for the unique than the non-unique nonwords (unique: 10.4%; non-unique: 8.3%), but this difference was not significant [all Fs < 1].

There interaction between nonword type and trial type was not reliable [$F_1(1,19) = 3.43$, MSE = 5654.83, $p > 0.08$; F_2 and min F' < 1]. The *input-switch* account had predicted such an interaction (see the Introduction to this chapter).

As in the word data (i.e., the interaction between language and frequency), there was a suggestion in the nonword data that bilinguals were more sure of German than of English. They were faster (by 6ms) and more accurate (by 7.4%) on German than on English trials (RT: significant by subjects only; errors: at least $p < 0.05$ by subjects, items and min F'). An interaction between language and nonword type (significant only by subjects though) qualifies this finding. Individuals were faster and more accurate on German than on English unique nonwords. On non-unique nonwords they were slower but more accurate on German than on English trials (possibly allowing more time to search on a German trial, feeling they had a greater chance of identifying obscure German words). See Appendix 3 for the relevant analytical details.

No other interactions were significant.

In sum, there was no statistical support for a switch-cost on nonwords (but see Experiment 2), and little evidence of a nonword type effect (replicated by Experiment 2).

2.4 Summary of Experiment 1a

Overall, consistent with all four accounts (*external, internal, input-switch* and *associative connections*), Experiment 1a confirmed a significant cost for switching between two languages in a language-specific LDT, both in the overall word analysis and in the analysis based on a subset of high frequency words. For nonwords, the difference between switch and non-switch trials was statistically not significant. There was some limited evidence of a nonword type effect. Its data pattern (Table 2.4: a difference between unique and non-unique nonwords only on switch trials) suggests that nonword stimuli such as the ones used here may not be able to elicit cross-language effects, and that nonword type effects for unique/non-unique nonwords may reflect something else (perhaps a response strategy).

The results are consistent with the *external* account's proposition that the language-switch cost reflects the time required to shift between different LD task schemata (e.g., "accept if it is an English word" vs. "accept if it is a German word"). However, they can also be accommodated by the *internal* account's notion of activation differentials within the lexico-semantic system.

Experiment 1b: language-general LDT

2.5 Introduction

Experiment 1b varied the lexical decision instructions, while presenting the same stimuli and also using the alternating runs paradigm. It required participants to decide whether or not a letter string was a word - regardless of language. Such a task should alter the nature of the task schema. With only weak inhibition between the two task schemata (or two jointly active schemata), there should still be a language-switch cost, but it should be smaller than Experiment 1a's. For nonwords, however, the *external* account predicted the cost to disappear, given that it is possible for a nonword to be rejected without having switched schemata, and that there is likely to be little language information (language-change signals) feeding into the decision process paralexically. The *internal* account likewise predicts a smaller language-switch cost in a language-general task. It does so on the basis that under language-general instructions the inhibition from the language-node to the non-target lexical representations will be weaker, entailing less time to overcome this inhibition. On this account there should be a (small) cost for both words *and* nonwords. For words, the *input-switch* account predicts a *greater* language-switch cost in a language-general compared to a language-specific task. There being no cue to guide the input-switch (no screen colour changes, orthographically neutral words), the last-used switch setting is likely to be used again. This entails that the last-used lexicon will be searched, upon failure to find the item, the switch will be re-set and the other lexicon will be searched. I.e., on switch trials there are two search times (as well as one re-set time), compared to just one search time and one re-set time in the language-specific task. For word trials, the *associative connections* account also predicts greater delay on language-switch trials in the language-general compared to the language-specific task. This is because in the latter task colour cues could boost the activation of the lexical representations of the language to be switched into, aiding the switching process. In the language-general task, in contrast, there is no such help on a language-switch trial.

2.6 Method

2.6.1 Participants

20 German/English bilinguals (8 males, 12 females), drawn from the student population of UCL, participated. Ten had previously participated in Experiment 1a ('old') and ten

participated in Experiment 1b only ('new'). The latter had no previous experience with LDTs. Participants' age range was 22 to 34 years (mean = 28.4, *SD* = 3.6). All had normal or corrected vision. Participation was voluntary and remunerated at the standard departmental rate.

Language proficiency:

All (except for one) grew up with German and learned English at school as their second language at a mean age of 11.2 years (5.1). At the time of the experiment the participants had on average been using English continuously for 5.7 years (5.8), ranging between 1 and 25 years. This figure includes the one participant who grew up with English. Note that this person's experimental data did not differ from that of the others. For English the results of the LLEX test ranged 82-98 (mean = 91.3, *SD* = 4.9). For German the range was 72-99 (mean = 90.2, *SD* = 7.6), suggesting that this group too was comprised of reasonably balanced bilinguals. Again, however, the self-rating scores (both languages to add up to a combined total of 100 points) indicated a higher estimate for German proficiency (mean = 56.0, *SD* = 11.8) as compared to English. Participants reported their current use of the two languages to be predominantly English rather than German (again measured out of a combined total of 100 points). The mean for English was 73.0, *SD* = 22.7. Of these participants, 9 chose the German version of the questionnaire, 11 the English one. The ten 'old' participants and the ten 'new' ones were comparable in terms of the language proficiency measures.

2.6.2 Design

Analytically, Experiment 1b had the same design as Experiment 1a, with the addition of the between-subjects factor Group ('old'/'new'). This made it possible to ensure that any reduction in switch-cost (that is being predicted for the language-general task compared to the language-specific one) is not related to either practice effects nor to individual differences. Note: Experiment 1b was run 6 months after Experiment 1a, so that while the 'old' participants had once experienced the routine required of them, the second experimental session was not an immediate repetition of the first one, which could have made the two groups very disparate. Furthermore, 'old' participants were not shown the same stimuli as before and so had no practice of individual items. There was a balance of 'old' and 'new' participants for each set of stimuli (sets A / B).

2.6.3 Materials, Apparatus & Procedure

The materials were identical to those of Experiment 1a (see Tables 2.1 and 2.2). The apparatus and procedure were identical to Experiment 1a, except that there was *no* change of screen colour (it was black throughout), and that individuals were instructed to decide whether or not a letter string constituted a word *irrespective* of language (language-general LDT).

2.7 Results & Discussion of Experiment 1b

Results were analysed by mixed-factor repeated-measures ANOVA - for words and nonwords separately. For the word analyses (by subjects) the between-subjects factor was Group ('old' / 'new'), the within-subjects factors were Trial Type (language-switch / language-non-switch), Language (English / German) and Frequency (HF/LF). For the nonword analyses (by subjects) the factor Frequency was replaced with the factor Nonword Type (non-unique / unique), all other factors were the same. For the item analyses, Language, Frequency and Nonword Type were between-subjects factors, all other factors were within-subjects.

2.7.1 Data treatment

Errors (for words: 9.9%, for nonwords: 7.0%, overall: 8.5%) and outliers (for words: 2.4%, for nonwords: 4.1%, overall: 3.3%) were removed from the mean RT data before analysis. (Median analyses again had only errors removed. As median analyses follow the same pattern as the means analyses, only the latter are reported here.) Lower (and varying) degrees of freedom in the item analyses are due to removal of errors and outliers, which reduced some cells to zero. Error analyses were on arcsine transformed error proportions. Inspection of RTs and error rates indicated that there was no speed/accuracy trade-off. Mean values presented in the text and tables are all mean of means (of the subject data). The 'switch-cost' is the difference between switch and non-switch trial data.

2.7.2 RT and error analyses for word trials

The mean correct reaction times, standard deviations and percentages of errors (for words) are displayed in Table 2.5. It shows slower RTs on switch than on non-switch trials, but clearly this difference is smaller than that in the language-specific task. On average, the switch-cost for German appears to be larger than that for English, but this relates to the comparatively faster RTs for German LF items on non-switch trials (compared to the English LF items on non-switch trials). This language difference presumably reflects differences in knowledge. Analytical details of these and other effects are provided below.

Table 2.5

Experiment 1b (language-general lexical decisions). Mean correct RT (ms), standard deviations (in brackets) and error percentages per cell for all words, as a function of language (English/German), frequency (high/low) and trial type (switch/non-switch). The switch-cost (ms) is included to facilitate reading the table.

		Trial Type						
			Switch			**Non-Switch**		**Switch**
Language	**Word Type**	**Mean RT**	**(*SD*)**	**% Error**	**Mean RT**	**(*SD*)**	**% Error**	**Cost (ms)**
English	**HF**	645	(89)	2.3	631	(77)	2.3	14
	LF	803	(110)	27.5	799	(111)	25.8	4
German	**HF**	626	(82)	3.5	599	(61)	1.3	27
	LF	770	(168)	12.3	662	(86)	13.5	108

The two groups of participants (those who had previously participated in Experiment 1a vs. those who had not) did not differ statistically in their overall reaction times [group: Fs < 1]. 'Old' participants made fewer errors than 'new' ones (9.2% vs. 12.6%), but this was significant only by subjects [$F_1(1,18) = 15.20$, MSE = 0.02, $p < 0.001$; by items $p > 0.2$; min F' < 1]. Most importantly, there were no significant interactions between group and trial type either in the RT or the error analyses [F < 1 or $p > 0.2$]. This indicates that if the predicted reduction in switch-cost (from language-specific to language-general) is shown, then it is unlikely to be attributable to sampling differences or differences in experience. The data also suggest that either there is little effect of previous experience with a language-specific task, when carrying out a subsequent language-general task, or that the time interval (6 months) between the two was sufficiently long as to have eliminated any practice effects which might have arisen had the interval been shorter.

As in Experiment 1a, there were significant main effects of trial type, language and frequency in the RT analyses. The most important results here are the switch-cost data. Language and frequency effects (and their interaction) are subsidiary. The information for these effects is kept to a minimum here. The analytical details are in Appendix 4.

Reaction times on switch trials were 38ms slower than on non-switch trials (711ms vs. 673ms) [$F_1(1,18) = 10.06$, MSE = 2950.97, $p < 0.01$; $F_2(1,291) = 8.14$, MSE = 19497.89, $p < 0.01$; min F'(1,82) = 4.50, $p < 0.05$]. This suggests that language continued to play a role even though the task did not require it. Consistent with the *external* and *internal* accounts, but contrary to the *input-switch* and *associative connection* accounts, the switch-cost found here is clearly much smaller than that observed in Experiment 1a (there the difference was 117ms, here it was 38ms), indicating that task instructions have implications for performance. But this shall be investigated in more detail later on (see section 2.9*)*. In contrast to the results of Experiment 1a, there was no difference in error rate between switch and non-switch trials - 11.4% and 10.7% respectively [$F_1(1,18) = 2.06$, MSE = 0.01, $p > 0.1$; $F_2(1,316) = 1.15$, MSE = 0.18, $p > 0.2$; min F' < 1].

The following effects (language and frequency) are dealt with briefly. Analytical details are in Appendix 4.

As in Experiment 1a, RTs were faster (by 55ms on average) to German words than to English words [at least $p < 0.025$]. There were fewer errors for German words (7.7%) than for English words (14.5%). This difference was most marked on the item analysis [$p < 0.025$; marginal by subjects and on min F'].

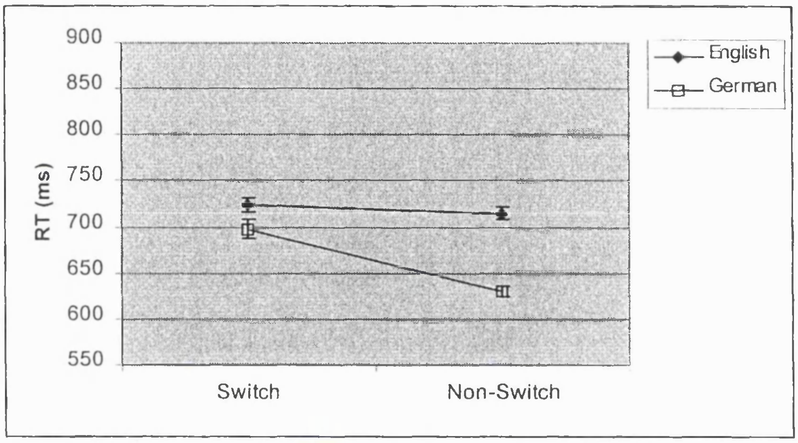
Responses to HF words were faster (by 134ms) than those to LF words [all $p < 0.001$] and there were fewer errors on HF words (2.4%) than LF words (19.8%) [all $p < 0.001$].

As in Experiment 1a, there was a significant interaction in the RT analysis between language and frequency [at least $p < 0.05$]. Although RTs were similar for HF words (German: 613ms; English: 638ms), RTs were much slower for English LF words (801ms) than for German LF words (716ms). The error data reveal the corresponding pattern. For HF words, the error rate was comparable for German and for English words (2.4% and 2.3%, respectively). In contrast, for LF words, the error rate was higher for English words (26.7%) than for German words (12.9%) [at least $p < 0.025$]. Such results suggest differential proficiency in English and in German, which is consistent with the self-ratings given by participants.

There was an indication of an interaction between trial type and language. However, this was significant only by subjects [$p < 0.025$; by items $p = 0.081$; on min F' $p > 0.2$]. The average switch-cost for German was 67ms (switch trial: 698ms, non-switch trial: 631ms), for English it was 9ms (switch trial: 724ms, non-switch trial: 715ms). As such, this appears at first to be an asymmetric switch-cost, with the larger cost for switching into the L1 (as found by Meuter & Allport, 1999, and Thomas & Allport, 2000). However, as can be seen in Table 2.5 and Figure 2.3, this relates to the much faster RTs for German LF words on non-switch trials (compared to their English LF counterparts). This, presumably, reflects differential knowledge in the two languages. In terms of errors there was no significant interaction between trial type and language [all Fs < 1].

Figure 2.3

Mean correct RT (ms) of words (averaged over HF and LF) on switch and non-switch trials as a function of language (German [L1] and English [L2]) in the language-general LDT (Experiment 1b). Error bars are standard error of the mean.



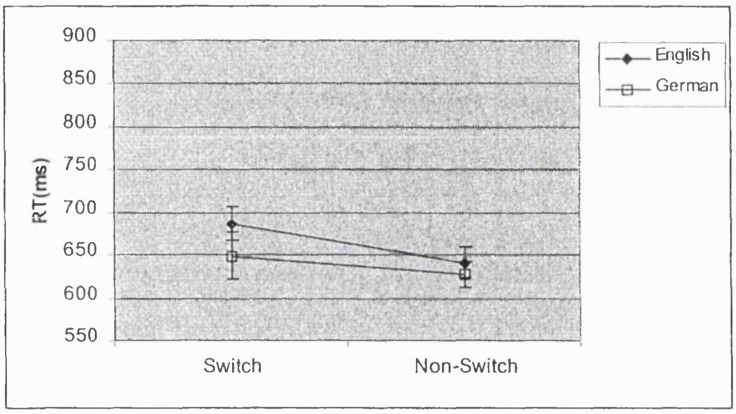
2.7.3 RT analysis for a subset of high frequency word trials

Since errors averaged only 2.7% on the critical HF set, no error analyses could be carried out. Note, the number of instances of HF words following HF words were exactly the same as in Experiment 1a. Thus again the data can only indicate a pattern, they do not permit strong claims. The RT pattern of this subset of materials was substantially the same as the overall analysis. Again, there was no significant effect of group [$F_1 < 1$]. There was, overall, a 34ms cost of switching between languages [$F_1(1,18) = 7.09$, MSE = 3213.45, $p < 0.025$]. This is in accord with the *external* account's proposition of weak mutual inhibition between the schemata (or of jointly active schemata - or even a single schema in conjunction with language signals affecting response competition paralexically). As could be expected on the basis of the overall analysis (similar RTs for English and German HF trials), there was no effect of language for these high frequency materials [$F_1(1,18)= 2.56$, MSE = 5237.12, $p = 0.13$]. There was *no* interaction between language and trial type in this subset analysis [$F_1 <$

1], with a 22ms difference between switch and non-switch trials for German words (649ms and 627ms, respectively) and a 46ms difference between switch and non-switch trials for English words (687ms and 641ms, respectively) – see Figure 2.4. This supports the interpretation of the combined HF and LF word data, that the interaction between language and trial type was not, in fact, an asymmetric switch-cost, but was an effect related to how long it took bilinguals to respond to LF English words.

Figure 2.4

Mean correct RT (ms) of the HF subset on switch and non-switch trials as a function of language (German [L1] and English [L2]) in the language-general LDT (Experiment 1b). Error bars are standard error of the mean.



2.7.4 Reaction time and error analyses for nonword trials

Data on effects that pertain perhaps more to participants' language background than to effects of language-switching are dealt with only briefly (for details see Appendix 4).

On average, 'old' participants were faster to reach a nonword decision (813ms) than 'new' ones (859ms), an effect significant only by items though [$F_1 < 1$; $F_2(1,287) = 44.33$, MSE = 17415.54, $p < 0.001$; min F' < 1]. 'Old' participants also made fewer errors (4.2%) than 'new' ones (9.1%), though by subjects and on min F' this was only marginally significant [$F_1(1,18) = 4.08$, MSE = 0.08, $p = 0.059$; $F_2(1,316) = 46.94$, MSE = 0.05, $p < 0.001$; min F'(1,21) = 3.75, $p = 0.066$]. Prior performance appears to have led to some learning in the task and to a change in response criterion (for example shorter decision deadlines for nonwords). But such a difference could also be attributable to some sampling differences. 'Group' did not interact significantly with any of the other factors.

There was no reliable effect of language switching for nonword trials (see Table 2.6), on either RTs or errors [all Fs < 1]. The average difference between switch and non-switch trials was 13ms in RT and 1.3% in errors. This is in accord with the *external* account's proposal that a nonword could be correctly rejected by either task schema, and/or that there is little language information affecting the response-competition paralexically. The *internal* account predicted a switch-cost for nonword stimuli, just as there was for words. (Indeed, the language-general task of Experiment 2, with greater power, also shows no evidence of a switch-cost for nonwords in a language-general task.)

The data suggest, as in Experiment 1a, that it took longer to reject non-unique nonwords (809ms) than unique ones (786ms), but statistically this effect was significant only by items [$F_1(1,18) = 3.93$, MSE = 11412.12, $p = 0.063$; $F_2(1,287) = 4.42$, MSE = 33607.55, $p < 0.05$; min $F'(1,61) = 2.08$, $p = 0.15$]. Individuals were more error prone for non-unique nonwords (8.0%) than unique ones (5.5%), reliably significant only by subjects and items [$F_1(1,18) = 8.30$, MSE = 0.02, $p < 0.01$; $F_2(1,316) = 4.04$, MSE = 0.08, $p < 0.05$; min $F'(1,135) = 2.72$, $p = 0.1$].

The *input-switch* account predicted an interaction between trial type and nonword type: no difference between non-unique nonwords' RTs on switch and non-switch trials (both trials involve an input-switch re-set, as discussed in the *Introduction* to this chapter) but faster RTs for unique nonwords on non-switch than on switch trials (in the latter case the input-switch has to be re-set, in the former it does not). The data in Table 2.6 suggest that this pattern may be present, but statistically there is no support for this interaction [$p > 0.05$ or $F < 1$].

Table 2.6

Experiment 1b (language-general LDT). Mean correct RT (ms), standard deviations (in brackets) and error percentages per cell for nonwords, as a function of language (English/German), nonword type (non-unique[1]/unique[2]) and trial type (switch/non-switch). The language-switch cost (ms) is included to facilitate reading the table.

		Trial Type						
		Switch			**Non-Switch**			**Switch**
Language	**Nonword Type**	**Mean RT**	**(*SD*)**	**% Error**	**Mean RT**	**(*SD*)**	**% Error**	**Cost (ms)**
English	**Non-Unique**	799	(110)	6.8	790	(105)	6.3	9
	Unique	753	(101)	5.8	784	(104)	7.8	-31
German	**Non-Unique**	825	(100)	8.8	823	(101)	10.0	2
	Unique	786	(97)	3.0	820	(100)	5.3	-34

[1] orthographically legal in either language; [2] orthographically legal only in the target language;

Overall, the main effect of language (RTs) was not significant [by subjects and items $p > 0.1$, min F' < 1], nor on error rates [all Fs < 1]. For errors there was an indication of language interacting with nonword type, but this was only significant by subjects, marginal by items and n.s. on min F' [by subjects $p < 0.01$; by items $p = 0.083$; min F' $p > 0.1$]. In this language-general task there were more errors on stimuli designated 'non-unique' on German trials (9.4%) compared to unique German nonwords (4.2%). There was little difference in error rates as a function of uniqueness for the 'English' nonwords (6.6% and 6.8% for non-unique and unique respectively). Conceivably, these stimuli differ in their overall similarity to words in the person's lexicon. For RTs there was no such interaction [all Fs < 1].

2.8 Summary of Experiment 1b

When the task was language-general, the cost for switching languages was smaller than when the task was language-specific. This is consistent with the *external* and *internal* accounts, but not the *input-switch* and *associative connections* accounts.

For words, the language-general task still showed a language-switch cost, despite the fact that the task did not *require* the individual to differentiate between the two languages. This could be accounted for, as proposed by the *external* account, by two schemata with weak mutual inhibition (or two jointly active schemata of which the selected one is temporarily more active, or even a single schema, with 'language-change' signals influencing response units paralexically). The comparison between the 'old' and 'new' participants showed very similar patterns of responses, so that any differences found between the two tasks (language-specific and language-general) can be attributed to the difference in task, rather than to sampling differences or practice effects. For nonwords there was no reliable switch-cost (as shall be seen, the more sensitive Experiment 2 did not show one either). This is in accord with the *external* account. The *internal* account, however, predicts a (small) switch-cost for nonwords, just as was found for words.

2.9 Comparisons between Experiments 1a and 1b

Two analyses of the overall RT data for words were performed in order to compare the switch-costs in the two lexical decision tasks. 1) A within-subjects analysis for the ten individuals acting in both experiments, and 2) a mixed-factor analysis for those acting in just one of the studies (ten in Experiment 1a and ten in 1b). These analyses used the same factors as before, with the addition of 'Task'. For clarity's sake, only the comparisons involving the nature of the decision task are presented. The main effects of language and frequency were substantially the same as in the previous analyses.

2.9.1 Task comparison for word trials (HF and LF together)

For the **within-subjects analysis**, RTs were overall 87ms faster in the language-general than in the language-specific task, but this effect was significant only in the item analysis [$F_1(1,9) = 3.99$, MSE = 44225.30, $p = 0.08$; $F_2(1,289) = 16.88$, MSE = 21382.84, $p < 0.001$; min $F'(1,14) = 3.23$, $p = 0.092$]. Participants taking part in both experiments reported finding the language-specific task more difficult than the language-general task.

Most importantly, as predicted by the *external* and *internal* accounts, the relative cost to RT of language-switching was significantly greater in the language-specific task than in the language-general one (for this set of participants it was on average 137ms and 26ms respectively) [task x trial type: $F_1(1,9) = 19.31$, MSE = 3133.48, $p < 0.01$; $F_2(1,289) = 8.26$, MSE = 29379.69, $p < 0.01$; min $F'(1,86) = 5.79$, $p < 0.025$].

The results of the **between-subjects analysis** substantially paralleled the within-subjects results above. RTs were on average 40ms faster in the language-general than in the language-specific task, but this was again significant only in the item analysis [$F_1(1,18) = 2.41$, MSE = 181063.01, $p = 0.13$; $F_2(1,279) = 131.64$, MSE = 20796.42, $p < 0.001$; min $F'(1,19) = 2.37$, $p = 0.14$].

Most importantly, consistent with the *external* and *internal* accounts, the language-switch cost was greater in the language-specific than in the language-general task (for this set of participants it was on average 90ms and 34ms respectively) [task x trial type: $F_1(1,18) = 16.57$, MSE = 10332.29, $p = 0.001$; $F_2(1,279) = 29.06$, MSE = 26730.37, $p < 0.001$; min $F'(1,44) = 10.55$, $p < 0.01$]. It is possible that the small group size was insufficient to establish the interaction against background noise.

The data of these task comparisons support the *external* and *internal* accounts, but go against the predictions made by the *input-switch* and the *associative connections* accounts, which had predicted a larger switch-cost in the language-general than in the language-specific task. This was not found.

2.9.2 Task comparison for HF subset trials

For both the within-subjects and the between-subjects analysis, the main effect of task was not significant [$p > 0.05$], but the main effect of trial type was significant [$p < 0.001$]. There was a significant task by trial type interaction for the HF subset [within-subjects analysis: $F_1(1,9) = 26.05$ MSE = 3159.95, $p < 0.001$; between-subjects analysis: $F_1(1,18) = 21.42$, MSE = 8108.42, $p < 0.001$]. For the within-subjects group the switch-cost was 158ms in the language-specific task, compared with 30ms in the language-general task. For the between-subjects group the switch-cost was 224ms in the language-specific task, compared with 37ms in the language-general task. There were no other significant effects [Fs < 1 or $p > 0.2$].

2.9.3 Task comparison for nonword trials

For nonword trials, the main effect of task was not reliably significant either within- or between-subjects. (Between-subjects: RT: n.s. by subjects and min F', but significant by items [$F_2(1,263) = 75.78$, MSE = 29413.72, $p < 0.001$]. Errors: Fs < 1 or $p > 0.2$. Within-subjects: RT: significant only by subjects [$F_1(1,9) = 6.07$, MSE = 56887.42, $p < 0.05$]. Errors: *p*s > 0.2.)

The main effect of trial type for RTs was significant between-subjects [$F_1(1,18) = 7.72$, MSE = 6552.31, $p < 0.025$; $F_2(1,263) = 14.44$, MSE = 39885.04, $p < 0.001$; min $F'(1,42) = 5.03$, $p < 0.05$]. For errors it was n.s. [all Fs < 1]. Within-subjects, for RT, this effect was significant only by subjects (n.s. by items and min F') [$F_1(1,9) = 11.74$, MSE =

10439.55, $p < 0.01$; by items $p > 0.2$; min F' < 1]. For errors it was n.s. [Fs < 1].

The main effect of nonword type for RT was significant by subjects and items in the between-subjects analysis (though marginal on min F') [$F_1(1,18) = 6.42$, MSE = 16298.45, $p < 0.025$; $F_2(1,263) = 8.04$, MSE = 38094.26, $p < 0.01$; min F'(1,56) = 3.57, $p = 0.064$]. For errors there was no effect [$p > 0.1$ or F < 1]. Within-subjects, nonword type was significant only by subjects [$F_1(1,9) = 5.10$, MSE = 8798.54, $p = 0.05$; by items and min F': F < 1] For errors there was no significant effect [Fs < 1].

Most importantly, as predicted by both the external and the internal accounts, the cost for language switching on nonword trials was smaller in the language-general task (average switch-cost: 8ms and 0ms for between-subjects and within-subjects respectively) than in the language-specific task (average switch-cost: 79ms and 111ms for between-subjects and within-subjects respectively). In the between-subjects case this task x trial type interaction was significant by subjects and items and marginal on min F' [$F_1(1,18) = 11.72$, MSE = 6552.31, $p < 0.01$; $F_2(1,263) = 46.86$, MSE = 27104.66, $p < 0.001$; min F'(1,28) = 9.38, $p < 0.01$]. In the within-subjects analysis this same interaction was significant only by subjects [$F_1(1,9) = 44.30$, MSE = 2789.62, $p < 0.001$; by items and min F' $p > 0.2$]. Presumably the small participant numbers were insufficient to establish an effect against the noise of the data.

As predicted by the external account, the RT pattern to unique and non-unique nonwords was not significantly in the language-general and the language-specific task [nonword type x task: F < 1 or $p > 0.14$]. This held true for both the within-subjects and the between-subjects comparison. For errors there was a marginal interaction between nonword type and task [within-subjects: by subjects $p = 0.084$ but $p > 0.2$ by items; between-subjects: by subjects $p = 0.068$, by items $p = 0.067$][45]. As all these analyses were marginal, and none were significant on min F' [Fs < 1], these data are not considered further here. If the internal account is the correct model, then this finding (no reliably significant nonword type x task interaction) could mean that the nonwords here did not elicit cross-language influences. If they had, then, on the internal account, the nonword type effect should have been smaller in the language-specific than in the language-general task. (In the language-specific task, strong suppression of the non-target language lessens the influence from the non-target language.) Contrary to the findings, the input-switch account had predicted a greater difference between

---

[45] Within-subjects: there were 4.4% more errors on unique than non-unique nonwords in the language-specific task, but 1.1% fewer errors on unique than non-unique nonwords in the language-general task. Between-subjects: there were 0.2% more errors on non-unique nonwords in the language-specific task, and 3.7% more errors on unique than non-unique nonwords in the language-general task.

unique and non-unique nonwords in the language-general than in the language-specific task[46].

2.10 General Discussion

These experiments explored the effects of language-switching on performance (RT and errors) in two versions of the lexical decision task. In Experiment 1a German/English bilinguals made language-specific lexical decisions. They decided if a letter string was either a real word in English or a real word in German. On alternate trials, cued by a change of screen colour, the language (in which lexical decisions were to be taken) changed. In Experiment 1b bilinguals decided whether the letter string was a word or not, irrespective of language (language-general LDT). Language still switched on alternate trials, but this fact was not 'advertised' to participants.

There were two main findings: a) On word trials (both for the overall word set and for the HF subset) individuals took longer to reach a lexical decision on language-switch trials than on non-switch trials I.e., there was a language-switch cost. This result confirms, for a new group of bilinguals, a finding reported by Thomas & Allport (1995). b) The switch-cost for words (both in case of the full set and the HF subset) was reduced, but not abolished, when individuals were required to perform a language-general LDT. This is consistent with the external and internal accounts, but not the input-switch and associative connections accounts. For nonwords the data patterned in a similar way, but the statistics did not always confirm the patterns.

In the following these effects are considered in more detail.

2.10.1 Processes involved in carrying out the tasks

For the language-specific lexical decision task, individuals must determine whether or not the letter string is a word in the designated language. Following Green (1995) it is supposed that there are two competing (mutually inhibitory) lexical decision schemata. The schemata link the output of the lexico-semantic system to the response mechanism. Cued to decide whether

---

[6] Recall, there being neither colour nor orthographic cues for the non-unique nonwords, such items required searching two lexica and re-setting the input-switch in the language-general task. In the language-specific task, colour cues could be used to guide the required setting of the input-switch immediately, and so avoid searching both lexica. That is, non-unique nonwords should be overall slower in the language-general than in the anguage-specific task. Unique nonwords provide a cue for the input-switch in both the language-general and he language-specific LDT, so that for them overall RTs could be similar in the two tasks. In sum, the difference between unique and non-unique nonwords should be exaggerated in the language-general task.

the letter string is a word in English, the activation of the lexical decision schema for English increases (and inhibits the one for German). On a following non-switch trial, the English schema is still dominant, allowing a more rapid response. On a switch trial, the German schema must come to dominate and must therefore overcome its suppression. With strong mutual inhibition it takes a significant amount of time to achieve this. Given non-selective access, a switch-cost arises because of response factors.

Experiment 1b changed the lexical decision task so that bilinguals could respond on the basis that the letter string was a word regardless of its language. In such a task there is no target language and hence no requirement to a) suppress the activity of the non-target system (internal account) or b) to establish schemata with strong mutual inhibition (external account). In principle, individuals could give a 'yes' response as soon as a matching entry has been found in the lexico-semantic system, without reference to language information. However, there was still a switch-cost for real words. Participants apparently still differentiated between languages. The data are in accord with the proposal that individuals established task schemata with weak mutual inhibition, or two jointly active schemata, or even a single schema, with language signals feeding into the response mechanism paralexically. The question whether in such a language-general task bilinguals establish just a single task schema, is discussed further in Chapters 3 and 6.

Overall, the data pattern supports the external account best. The internal account predicted a (small) switch-cost for nonwords in the language-general task, for which there was no statistical evidence, but otherwise the internal account was able to accommodate the data patterns. The input-switch and the associative connections accounts were unable to predict the data of most importance: the switch-cost reduction from the language-specific to the language-general task.

Could individuals in a regular switching paradigm not prepare for a language-switch in advance? There was, after all, a one second interval between trials. If preparation occurs, the results indicate that it is a process which is not completed. The data presented here are in accord with the notion to be found in the literature that preparations can be made up to a point, but that ultimately the response is driven by the incoming stimulus (e.g., Rogers & Monsell, 1995; Meiran, 1996; Monsell, 1996; Meuter & Allport, 1999). On the basis of task instruction (i.e., using an executive control process), individuals establish a control structure for the task (appropriate task schemata with strong mutual inhibition / task schemata with weak mutual inhibition or two jointly active schemata) and then allow the stimulus input to drive the response ('exogenous control', Monsell, 1996).

2.10.2 Asymmetric language-switch costs

In this study there was no evidence of 'asymmetric switch-costs', as was found, for example, by Thomas & Allport (2000, Experiment 2). They found greater switch-costs when switching into L1 than when switching into L2. It is proposed here that the difference between the two studies (the present one and Thomas & Allport's) is related to stimulus list characteristics. In Thomas & Allport's (2000) Experiment 2 'wrong language words' were shown (e.g., a French word on a trial designated as 'English'). Participants must reject such items, even though they are words. Assuming that language signals (or other signals emerging from a specific lexicon) play a role in task schema selection, it is possible that, for an imbalanced bilingual (as Thomas & Allport, 2000, state their participants were), L1 signals (or other signals from the L1 lexicon) are better able to trigger L1 task schemata, compared to L2 signals (or other signals from the L2 lexicon) triggering L2 task schemata. If this is the case, then, in the presence of wrong language words, individuals will need to strongly inhibit an L1 task schema during the time that an L2 schema is required. Otherwise there would be a great chance for false positive replies to L1 words presented on L2 trials. In contrast, an L2 schema need not be as strongly inhibited during the time that an L1 task schema is required. Under such circumstances, more disinhibition is required for shifting to an L1 task schema than there is for shifting to an L2 task schema. When there are no 'wrong language words' and bilinguals are more balanced, one may presume that inhibition between the two languages' task schemata is comparable, resulting in symmetric rather than asymmetric switch-costs. This issue is discussed in more detail in Chapter 6.

2.10.3 Nonword data patterns

Taking a non-selective access view, one might have expected to find a nonword type effect (RT unique < RT non-unique). Unique nonwords have little or no means of eliciting activation in the 'other' language, since, due to their orthography, they are unlikely to have many (if any) real-word neighbours in the 'other' language (and consequently little cross-language competition). Non-unique nonwords, in contrast, are likely to have such neighbours, and so should elicit activation in both language systems (and consequently comparatively more cross-language competition). If this is the case, then should there not be evidence of cross-language influences? This is what Altenberg & Cairns (1983) suggested they found. But, as argued in the introduction to this chapter, it is unclear whether such a finding (RT unique < RT non-unique) reflects cross-language influences (recall, Altenberg & Cairns found this pattern for unilinguals as well as bilinguals, for the former it was no stronger than for the latter, and for the bilinguals it was absent in their weaker L2). Instead, it is possible that the nonword type effect Altenberg & Cairns found was an orthographic response strategy, which allowed one type of nonword to be rejected more quickly than the other. It may be that this type of nonword stimulus is not suitable for finding cross-language influences. In any event, not finding cross-language influences does not invalidate the results, which provide evidence regarding the origin of language-switch costs. The question of cross-language influences shall be returned to in Chapter 5, where words that have direct representations in both lexica (interlingual homographs) will show that such influences exist (contrary to what the selective access input-switch account would predict). The issue of cross-language influences is further addressed in Chapter 5, and discussed in more detail in Chapter 6.

2.11 Conclusion

The experiments presented in this chapter are consistent with the hypothesis that the cost for switching between languages is attributable to processes outside the lexico-semantic system (external account). The suggestion made here is that it is (language) task schemata and their relationship to each other which gives rise to the language-switch cost and its modulation. The internal account is also able to predict the most important data patterns obtained. However, it predicted a remaining switch-cost for nonwords in the language-general task, which was not found. The input-switch and associative connections accounts are able to capture elements of the findings, but not the whole data pattern. In fact, they predict the opposite effect for the most important prediction made by the external account (switch-cost *reduction* in the language-general task, rather than an increase).

2.12 Outlook to Experiment 2

The inter-task comparisons in this experiment were based on small samples (in the within-subjects analysis, ten participants performed in both tasks; in the between-subjects analysis, ten participants performed only in Experiment 1a and ten performed only in Experiment 1b). Thus these comparisons were not as powerful as one would have wanted them to be. This is presumably the reason why some of the effects in the task comparison differ for the between- and the within-subjects analyses (e.g., section 2.9.3). Experiment 2 (Chapter 3) will contrast language-specific and language-general tasks fully within-subjects, using larger subject groups, to obtain greater analytical power.

In Experiment 1a (language-specific LDT) language was cued by screen colour. Thus, a 'general change' took place on language-switch trials (the screen colour changed), but not on language-non-switch trials (the screen colour remained unchanged). Furthermore, only in the language-specific task was there such a 'general change', since in the language-general task words were presented on a black screen throughout. It is possible that the trial type effect was confounded with 'general change' in Experiment 1a, as likewise the task effect could have been confounded with 'general change'. Experiment 2 also contrasts language-specific and language-general LDTs, but using a 'clocks' type presentation, (developed by Rogers & Monsell, 1995, for task-switching studies). In this paradigm, items are presented in one of four quadrants. For the language-specific task participants are told that particular squares are 'German', while the others are 'English'. In the language-general task participants are not told about any association of languages to particular squares. In this paradigm the movement from one square to another is the 'general change'. Thus for switch and non-switch trials alike, and for language-specific and language-general tasks alike, any influence exerted by 'general change' is comparable, avoiding any possible confound of 'general change'.

To further investigate the external account, the following points need to be borne in mind: a) Task schemata do not operate in a vacuum. They are presumably influenced by the various signals available during the processing involved in lexical decision tasks (signals that pertain to individual items when unique identification is achieved, familiarity or wordness signals, language signals, task cue signals, etc.). b) The experiments are carried out by people whose lives are spent changing and adapting, striving for optimal performance (maximal performance at minimal cost). Thus it is likely that individuals develop 'strategies' for responding, according to experience and the constraints of task type and stimulus list composition. These points pose questions regarding stimulus list composition and the order of carrying out the different tasks - both are addressed in the following chapter.

Finally, the data provided by Experiments 1a and 1b were unable to clearly distinguish between the internal and external accounts of control. The external account suggests that there may be no language-switch cost for nonwords in the language-general LDT (see Introduction to this chapter), while the internal account would predict such a cost (albeit a small one). No language-switch cost for nonwords was found in the language-general LDT, favouring the external account. However, given the uncertainty about how the nonwords used here affect processing (see footnote 39), it would seem unwise to place much weight on this point. As was said in the Introduction to this chapter, the word data is of primary importance here, and on the word data there was no means of differentiating between the two control accounts. The word data did, however, allow one to distinguish clearly between the two control accounts and the two early models (input-switch/associative connections). The role of Experiment 2, then, is to distinguish between the two control accounts more clearly. This is hoped to be achieved by refining the investigation as outlined above.

III

Experiment 2: The effects of task order and list composition on language-switch costs in language-specific and language-general lexical decision tasks

3.1 Introduction

Experiments 1a and 1b were, except for the question of the switch-cost on nonword trials in the language-general task, consistent with both the external and internal accounts of control. Experiment 2 tests these two accounts further, once again addressing language-switch costs in lexical decision tasks. The investigation is fine-tuned by two additional manipulations: *list composition* and *task order.* Where Experiments 1a and 1b presented only one list type (equal proportions of HF and LF words), Experiment 2 contrasts 'pure list' (all words are HF) and 'mixed list' (words are HF and LF) conditions. Where previously the task order was only language-specific before language-general, Experiment 2 balances task order over participants. Furthermore, Experiment 2 uses the 'clocks' paradigm (Rogers & Monsell, 1995) instead of the colour paradigm used in Experiments 1a/1b, to avoid the possible confound with 'general change' discussed at the end of Chapter 2. The present experiment is conceived in such a way as to allow analyses to focus on a critical stimulus set, consisting of HF words preceded by HF words (analogous to the 'HF subset' in Experiments 1a/1b). Analysing the variables of interest in this manner ensures a comparable data set for the different list conditions (i.e., the same HF words, under comparable local contexts, set within different list types). This method also ensures a maximum number of correct responses, both on the investigated trials themselves and on preceding trials.

As shall be discussed below, list, task and task order effects do not function in isolation, they interact. But to clarify the influence of each, they are first considered separately. Predictions for Experiment 2 made within this section are based on the external account. Predictions made by the internal account follow in a separate section.

List effects: Stringency of response criterion and its implications for paralexical effects of language and wordness signals. Individuals confronted with a 'pure-HF' list, are likely to adopt a more relaxed response criterion, allowing wordness signals to play a considerable role paralexically (i.e., allowing wordness signals to bias towards a 'word' response). However, language signals are likely to feed into the decision process even when a relaxed response criterion is adopted, since it is unlikely that a) bilinguals can ignore the fact that a word is a word in one language rather than another (as discussed in Chapter 2), and b) when nonwords are word-like, it is unlikely that lexical decisions would be taken solely on the basis of wordness (or familiarity). In contrast, as discussed in Chapter 1, individuals confronted with a 'mixed-frequency' stimulus list, are likely to use a more stringent response criterion (engage in more thorough processing). These individuals will presumably use wordness signals to some degree (as indicators to reply 'yes'), but they will need further information than that provided by the wordness signals, in order to confidently distinguish between LF words and word-like nonwords. That is, they will presumably not attach much weight to the paralexical wordness signals, since the proportions of cases, where this is a helpful cue, are insufficient (see section 1.6.6). Furthermore, given LF words and word-like nonwords, these participants are likely to make as much use as possible of any information available to them, including language signals. Apart from attending to language signals in order to help make correct lexical decisions, language signals are likely to become increasingly prominent when processing is more thorough[47].

The implications for language-switch costs, of placing more weight on wordness vs. language signals, are as follows. Following Thomas & Allport (2000), the assumption is made that perceiving a change of input-language biases bilingual participants to changing their response (see section 1.6.4). The more weight is placed on language signals, the greater the incorrect bias will be on language-switch/response-repetition trials. With comparatively more weight placed on language signals given a mixed stimulus list, one might expect a larger language-switch cost for a mixed than for a pure condition. However, the presence of word-like nonwords presumably forces even participants in a pure-HF condition to engage in relatively thorough processing, rather than basing responses solely on wordness signals. In

---

[47] Grainger & Dijkstra (1992, referring to unpublished data) showed lexical decisions to be made more quickly than language decisions (using the same stimuli for both tasks, the latter were on average 200ms slower). Grainger & Dijkstra (1992) explained this finding by saying that a language decision could be made once a language-node becomes active above a criterial amount. Given that language-nodes receive their activation from lexical representations, language information presumably lags behind wordness information, given a situation where single words need to be processed, and 'top-down' predictions are ruled out. On the notion of a language tag, one could also accommodate language decisions being made more slowly than lexical decisions. If a language tag is attached to a lemma representation, then that representation has first to be activated, before the tag attached to it can be taken as evidence that the item in question is a word of language X. In a language-decision task, paralexical decision processes using wordness or familiarity information will not help in deciding which response is correct. In contrast, making lexical decisions may be faster because wordness / familiarity information can bias towards the required response, rendering the execution of the response faster.

sum, a language-switch cost is expected for the 'pure-HF' condition just as for the 'mixed' condition, because in both cases processing is likely to be sufficiently thorough for language signals to become available and to influence processing.

Task effects: A language-specific task *requires* placing weight on language signals. Thus participants need to do so irrespective of what their stimulus list composition permits them (e.g., relaxed criterion given a 'pure-HF' stimulus list). In contrast, a language-general task does *not* require giving much weight to language signals; it permits placing more weight on wordness signals.

The combination of list and task effects: In a language-specific task, the task requires attending to (placing weight on) language signals. Given compliant participants, one may well expect language signals to play a significant role in the language-specific task, regardless of stimulus list composition. Accordingly, in this task there should be little difference between a pure and a mixed condition with regard to language-switch costs: The weight of language signals in a paralexical decision process should be similar. In contrast, in a language-general task, individuals confronted with a 'pure-HF' list can make use of this task's leniency, allowing them to place much weight on wordness signals. However, as argued above, participants are likely to be influenced by language signals to some degree. Accordingly, they should experience a language-switch cost, albeit a small one.

Individuals confronted with a 'mixed' stimulus list are likely to have a more stringent criterion for responding 'yes'. As a consequence of this more thorough processing, language signals are likely to become more prominent, leading to the expectation of a clear language-switch cost for the mixed condition in the language-general task. In short, in the language-general task, in both conditions, language signals are expected to play a role, but in the pure-HF condition bilinguals are likely to place overall comparatively less weight on them, and instead favour wordness signals. The implications of this difference become apparent when considering the impact of task order.

List, task and task order effects: Participants shown a 'pure-HF' list and completing the language-general task first ('language-general first'), will learn that placing weight on 'wordness' signals permits fast and correct responses. Having learned this, they are likely to continue to place weight on wordness signals in a following language-specific task (if there are no 'wrong language' words to engender errors), although language signals should begin to play more of a role now that task instructions require it. Individuals in a 'pure-HF' condition completing the language-specific LDT as their first task ('language-specific first'), will presumably simply adhere to the instructions given, and thus place more weight on language signals than on wordness signals. Given the proposed effect of 'language-has-changed' signals biasing towards the response previously not used, and that the critical HF words are all preceded by (HF) words, the prediction follows that, shown a 'pure-HF' list, the 'language-specific first' task should show a larger switch-cost than the 'language-specific second' task. In contrast, confronted with a 'mixed' list, individuals are likely to place less weight on wordness signals and more on language signals, even in a language-general task, and to continue in this manner in the following language-specific task. It follows that, shown a 'mixed' list, individuals should show comparable language-switch costs in 'language-specific first' and 'language-specific second'.

In sum, for Experiment 2, the external account predicts: 1) Given a 'pure-HF' stimulus list, a language-specific task carried out after a language-general one is likely to elicit a smaller language-switch cost than a language-specific task carried out without prior language-general experience. For a 'mixed' condition there should be little difference between a language-specific task that is 'first' and one that is 'second'. 2) Based on previous data ('old' and 'new' participants, Experiment 1b) it is predicted that language-general tasks should show comparable language-switch costs, whether they are carried out as 'first' or 'second' tasks. The basis for this prediction is that when permitted to employ a more lenient task schema set-up, bilinguals appear to do so; they do not continue to make language-specific decisions (with which they have previous experience) when the current task instructions allow language-general decisions.

What are the predictions made by an internal account?

Consider first *list composition.* Experiments 1a and 1b did not show a trial type by frequency interaction. This suggests that the activation from the lexical level to the language-node level is similar for HF and LF words. The prediction follows that the experience associated with 'pure-HF' and 'mixed' lists will not change the language-switch cost.

Consider now *task.* On the internal account the degree of inhibition (language-node level to lexical level) is assumed to be determined by the task instructions. Asking a participant to make language-specific lexical decisions, this individual could comply by choosing a setting of strong inhibition of the non-target language's lexical representations. In contrast, participants carrying out a language-general task would have a setting of weak inhibition of the non-target lexical representations. The stronger the inhibition, the greater the activation decrement which needs to be overcome on a language-switch trial. Accordingly, the language-switch cost is larger given language-specific than given language-general task instructions.

Consider now the combined effects of *task* and *task order*. If participants comply simply with task instructions (as outlined above), then a language-switch cost should be the same for a language-specific task carried out first and one carried out following a language-general task. Alternatively, participants might be influenced by previous task experience. When carrying out a language-general task, a setting of weak inhibition would be used and practised. When next carrying out a language-specific task, participants could recall the settings they employed before (weak inhibition of the non-target language), provided there are no 'wrong language' words to make such a strategy risky. In this case the language-switch cost would be smaller when the language-specific task follows the language-general one, compared to when it comes first. In short, participants could either allow their previous experience to primarily influence the settings they select, or they could allow task instructions to primarily affect the settings they select. The results of Experiment 1a (compared to Experiment 1b) suggest that task instructions play a prominent role in deciding what degree of inhibition to use (otherwise, participants in a language-specific task, not forced to make language-specific decisions given the absence of 'wrong language' words, could have made language-general decisions despite language-specific instructions).

Consider now *task, task order* and *list composition*. The external account used the notion of paralexical decision processes to argue why individuals in a 'pure-HF' condition (but not a mixed condition) should show an effect of previous task. Paralexical decision processes are not unique to the external account, they could certainly be at work given an internal account. Nonetheless, the internal account does not predict the same switch-cost reduction pattern as was proposed by the external account. Consider a language-switch trial. The previous inhibition of the now-target language (non-target on the previous trial), takes time to overcome, thus slowing the development of a general wordness signal which could bias towards the required 'yes' response. That is, the time taken to activate 'yes' sufficiently above 'no' is not shortened. Thus, even when much use could theoretically be made of wordness signals paralexically ('pure-HF' condition), such signals are not available early on, and therefore cannot have the proposed, helpful, biasing effect.

In sum, the internal account is unable to predict a switch-cost reduction ('language-specific second' < 'language-specific first') for only the 'pure-HF' condition. If previous task experience is of primary importance, then there should be a switch-cost reduction irrespective of condition (that is, both for 'pure-HF' and 'mixed-frequency' conditions). If task instructions are of primary importance, then there should be no switch-cost reduction - for any condition.

To test the two accounts (internal / external), Experiment 2 requires regular switching between languages (as in Experiments 1a and 1b). It compares LDT performance in 'pure-HF' and 'mixed' conditions. In the 'pure' condition all words were HF (the '100/0' condition). In one mixed condition real words were 50% HF and 50% LF (the '50/50' condition). In a second mixed condition real words were 75% HF and 25% LF (the 75/25 condition).

The latter mixed condition was used for a (subsidiary) investigation, namely whether there is simply a 'pure' / 'mixed' effect, or whether the actual proportion of HF to LF words gives rise to 'proportion effects'. For instance, individuals seeing many more HF words than LF words might place more weight on wordness signals than individuals who see an equal number of HF and LF words. Alternatively, and more probably, the presence of the LF words (and word-like nonwords) would enforce a stringent response criterion even when the stimulus list showed more HF than LF words. In short, the two mixed conditions should not differ from one another with regard to the lack of a task order effect in the language-specific task.

The same (word-like) nonwords were used in all conditions. Each participant carried out both a language-specific and a language-general task, so that the task effects could be investigated within-subjects. For half of the participants in each condition the language-specific task preceded the language-general one, for the other half this was reversed (i.e., order effects were between-subjects). Each participant was assigned to only one frequency condition. A more detailed account is provided in the Materials section.

Addressing frequency conditions between-subjects could be argued to be less sensitive than doing so within-subjects. However, the main aim of Experiment 2 was to investigate language-switch costs in the two LDTs and how they are influenced by task order, in order to differentiate between the external and internal accounts. Frequency condition was a necessary part of this, in that it can influence the response criterion, but it was not a factor pursued in its own right (contrary to the unilingual frequency blocking experiments by, for example, Glanzer & Ehrenreich, 1979, and Gordon, 1983).

As discussed in Chapter 2, the nonword data are unable to play a great role in distinguishing the two accounts, given the uncertainty of the processes involved. Therefore, the main focus is on the word data; nonword data shall be only briefly dealt with in this chapter's Results section.

3.1.1 Summary of predictions

(1) Replicating Experiments 1a and 1b, an effect of trial type was predicted (slower RT on language-switch than non-switch trials). It was also expected (by both external and internal accounts) that trial type and task would interact (greater language-switch cost in the language-specific than in the language-general task).

(2) Additionally to these replications of effects found in Experiments 1a and 1b, it was predicted by the external account that the language-switch cost would be reduced in a language-specific task that followed a language-general task (compared to the cost in a language-specific task when there was no previous language-general task experience), but that this would only be the case for the 100/0 condition. In the language-general task there should be no order effects. As discussed earlier, the internal account cannot predict a task order effect for just one task (language-specific) and just one condition (100/0); nor can the input-switch account[48].

(3) Neither of the two mixed conditions (50/50; 75/25) would show a reduced language-switch cost in 'language-specific second' compared to 'language-specific first'.

---

[48] If previous experience helped to speed up the process of re-setting the input-switch, then this should affect the language-specific and language-general tasks equally and all conditions equally.

3.2 Method

3.2.1 Participants

A total of 96 German/English bilinguals (46 male, 50 female; mean age = 27.0, *SD* = 4.5; range: 20-40), drawn from the student population of University College London, participated. None had previous experience with lexical decision tasks. All participants had normal or corrected vision. Participation was voluntary and remunerated at the standard departmental rate.

Language proficiency:

Participants grew up with German and started to learn English at school as their second language at a mean age of 10.7 (2.3) years. They had become fluent in English at a mean age of 18.5 (5.0) years and had on average been using English predominantly for the past 8.6 (5.6) years.

Following the experiment, participants carried out the LLEX tests (Meara, 1994 – for details see Chapter 2) and completed the language background questionnaire (see Appendix 1). In the LLEX tests, out of a maximum of 100, results ranged 72-99 for English (mean = 89.6, *SD* = 5.8), and 68-99 for German (mean = 91.9, *SD* = 5.8), suggesting that these bilinguals were reasonably well balanced. However, the result of the self-rating (self-assessing the relative proficiency in German compared to English, out of a combined total of 100), indicated somewhat higher estimates for German than for English (German: mean = 57.8, *SD* = 11.4). Participants reported the current use of their two languages (again out of a combined total of 100) to be predominantly English (mean = 69.0, *SD* = 20.2). 46 participants chose the English version of the questionnaire, 50 the German version.

Participants were randomly assigned to one of the three conditions (with the proviso of comparable numbers of males and females in each). The language scores for the participants per condition indicate that there were some small differences between groups, but that they were overall comparable:

a) the 100/0 condition:

Thirty-two German/English bilinguals (16 male, 16 female; mean age 27.8 (4.2) years, range 20-39) were assigned to this condition. Participants started to learn English as their second language at a mean age of 10.6 years (2.6). They had become fluent in English at a mean age of 20.2 (5.9) years and had been using English predominantly on average for the past 8.9 (6.3) years.

The LLEX test results (maximum score 100) ranged 79-99 for English (mean = 90.9, *SD* = 5.2), and 71-99 for German (mean = 90.9, *SD* = 6.2). The self-rating (out of a combined total of 100) indicated slightly higher estimates for German than for English (German: mean = 54.4, *SD* = 11.8). The current use of the two languages (out of a combined total of 100) was predominantly English (English: mean = 75.3, *SD* = 20.7). 17 participants chose the German, 15 the English version of the language background questionnaire.

b) the 50/50 condition:

Thirty-two German/English bilinguals (15 male, 17 female; mean age 27.3 (4.4) years, range 22-40) were assigned to this condition. Participants started to learn English as their second language at a mean age of 10.6 years (2.4). They had become fluent in English at a mean age of 18.9 (5.3) years and had been using English predominantly on average for the past 8.5 (5.4) years.

The LLEX test results (maximum score 100) ranged 79-98 for English (mean = 90.2, *SD* = 5.3) and 85-97 for German (mean = 92.8, *SD* = 3.2). The self-rating (out of a combined total of 100) indicated somewhat higher estimates for German than for English (German: mean = 60.4, *SD* = 11.4). The current use of the two languages (out of a combined total of 100) was predominantly English (English: mean = 70.0, *SD* = 22.8). 19 participants chose the German version of the questionnaire, 13 the English version.

c) the 75/25 condition:

Thirty-two German/English bilinguals (15 male, 17 female; mean age 26.0 (4.8) years, range 21-42) were assigned to this condition. Participants started to learn English as their second language at a mean age of 11.1 years (2.4). They had become fluent in English at a mean age of 18.1 (3.8) years and had been using English predominantly on average for the past 7.9 (4.9) years.

The LLEX test results (maximum score 100) ranged 72-97 for English (mean = 87.6, *SD* = 6.5), and 68-99 for German (mean = 92.0, *SD* = 7.2). The self-rating (out of a combined total 100) indicated somewhat higher estimates for German than for English (German: mean = 59.9, *SD* = 11.5). The current use of the two languages (out of a combined total of 100) was predominantly English (English: mean = 68.2, *SD* = 21.4). 15 participants chose the German, 17 the English version of the language background questionnaire.

3.2.2 Design

A mixed-factor design was used to investigate a critical set of HF words (embedded among filler items). Firstly, there was an 'overall' analysis (on the critical HF words), to provide a general overview of the data. For this there were three within-subjects factors: Language (German/English), Trial Type (language-switch/language-non-switch) and Task (language-specific/language-general). The between-subjects factors were Condition (100/0; 75/25; 50/50) and Task Order (first task / second task). The prediction of task order effects (reduced language-switch cost in the 'language-specific second' vis-à-vis the 'language-specific first' task, but only in the 100/0 condition), was assessed with a planned comparison contrasting the 100/0 condition with the mean of the 75/25 and 50/50 conditions. A second planned comparison contrasted the 75/25 condition with the 50/50 condition, assessing whether the actual proportion of LF to HF words modifies the task order effects. For these two planned comparisons all factors were as for the overall analysis, excepting the factor Condition, which had only two levels (100/0 vs. 75/25+50/50 in the first case; 75/25 vs. 50/50 in the second case).

The filler items were used for investigating further secondary issues (frequency effects and nonword type effects). For the frequency analysis (relevant only for the 50/50 and the 75/25 conditions) the within-subjects factors were the same as for the main analyses, with the additional within-subjects factor Frequency (HF/LF). In the case of the nonword analyses (performed on all three conditions) the within-subjects factor Frequency was replaced by the factor Nonword Type (non-unique/unique), all other factors were identical. Recall that 'unique' refers to nonwords orthographically legal only in one language (English or German, respectively); 'non-unique' refers to nonwords orthographically legal in both languages.

Trial sequences:

Language switches occurred on alternate trials (EEGG etc.). Following Rogers & Monsell (1995), items were presented using the 'clocks' method. Items appeared in a square (situated centrally on the screen) which was divided into four quadrants. Successive items appeared in different quadrants, moving around in a clock-wise fashion. For all participants and for both tasks, the first item of each experimental block appeared in the top left quadrant of the large square. For the language-specific LDT the 'starting language' was counterbalanced across participants in the three conditions and the two task order assignments. There were four possible starting sequences: EEGGEEGG etc.; EGGEEGG etc.; GGEEGGEE etc.;

GEEGGEE etc. Thus the two English squares could be either the two top, the two bottom, the two right or the two left squares; the opposite ones were the German ones.

In the language-general LDT stimuli were also shown in the quadrants, using the four different starting sequences as in the language-specific task (but without assignment of language to location). Since a quarter of the stimuli were non-unique nonwords, it was thought that participants should not become aware of any regularity of item position and language affiliation. Indeed, when asked following this task, individuals were surprised to be told that there had been such a pattern. Nonetheless, in order for the language-specific and the language-general tasks to be comparable, the 'starting language' was counterbalanced in the language-general task in exactly the same way as in the language-specific task.

Participants were not given the same presentation order of languages (for example, EEGG) in both their first and their second experiment. This was done as a further precaution to avoid that those participants, who were to carry out the language-general task after having previously completed the language-specific one, would notice the same language pattern they had experienced before. Had they made such a discovery, it is likely that the task would have ceased to be a language-general one, as language sequences would have become evident.

Half of the participants in each condition carried out the language-specific task first, followed a week later by the language-general LDT. The other half received the reverse order.

For each participant the starting sequence was identical in the practice block and subsequent experiment (e.g., if a participant saw the sequence EEGG in the practice block, they had the same sequence in the experiment). For participants in the 100/0 condition all words in the practice block were also HF – for the other conditions the practice block had 50/50 and 75/25 HF/LF proportions, matching their experimental conditions.

Words and nonwords alternated in such a way that there were not more than four 'yes' or 'no' responses consecutively. Participants were shown an equal number of word and nonword stimuli.

3.2.2.1 Design of the experimental blocks

Within each of the six experimental blocks there were some critical stimuli, embedded amongst filler items:

Critical stimuli:

Per experimental block (this applies to each of the three conditions) there were eight 'critical'

HF words: four English (two on switch and two on non-switch trials) and four German (two on switch and two on non-switch trials). All of these HF words were themselves preceded by HF words. Thus the local context of the critical items was held constant, while the overall list composition was manipulated.

General construction of experimental blocks:

Each experimental block consisted of a total of 64 experimental stimuli (this number includes the critical stimuli), preceded by one filler trial. Thus each participant saw a total of 390 stimuli during the experiment proper. The initial filler trials were included to ensure a clear designation of trial type for the first measured trial in each block (these filler trials were excluded from the analyses).

Within an experimental block each participant saw each type of stimulus (for English and German respectively: HF; LF; non-unique; unique) several times, the exact number of the real word stimulus types depending on the condition:

In the 100/0 condition: Per block each participant saw 16 German and 16 English HF stimuli (per language eight on switch, eight on non-switch trials). Non-unique and unique NWs were each seen eight times in each language (balanced over switch and non-switch trials). There were no LF words in this condition.

In the 50/50 condition: Per block each participant saw 8 German and 8 English HF stimuli (per language four on switch, four on non-switch trials), as well as 8 German and 8 English LF stimuli (per language four on switch, four on non-switch trials). Non-unique and unique NWs were each seen eight times in each language (balanced over switch and non-switch trials).

In the 75/25 condition: Per block each participant saw 12 German and 12 English HF stimuli (per language six on switch, six on non-switch trials), as well as 4 German and 4 English LF stimuli (per language two on switch, two on non-switch trials). Non-unique and unique NWs were seen eight times in each language (balanced over switch and non-switch trials).

The following constraints applied for all items: 1) Items seen on language-switch trials by one half of the participants were seen on language-non-switch trials by the other half. 2) No participant was shown the same item twice – in the same language or in translation. 3) Successive items were unrelated semantically and associatively to avoid priming effects. For each participant the stimulus order was unique.

The assignment of stimulus sets (A / B; see Materials below) was counterbalanced across conditions and tasks in such a way that each set was seen equally frequently in each condition (100/0; 50/50; 75/25), in each task (language-specific / language-general) and by each task order (language-specific first / language-general first). No participant saw the same set of stimuli twice.

3.2.3 Materials

a) Words:

Two sets of words (sets A and B) were constructed with a total of 192 words in each. In each set, half of the words were German and half were English (i.e., 96 of each language). For the 100/0 condition all words were HF (N = 96 per language). For the 50/50 condition half the words were HF (N = 48 per language) and half were LF (N = 48 per language). For the 75/25 condition 75% of the words were HF (N = 72 per language) and 25% was LF (N = 24 per language). The words were matched for letter and syllable length across the two languages and across sets. Within languages they were, as far as possible, matched for frequency across the two sets for each of the three list conditions (using Kučera & Francis, 1967, and Ruoff, 1990, for English and German respectively). The various constraints on the items available for selection necessitated including some items with particularly high frequencies, rendering the sets' mean values uneven. CELEX values (Baayen, Piepenbrock & van Rijn, 1993) were not available at the time of stimulus assembly, but have been included here for comparison. See Table 3.1 for the mean parameter values of the critical HF set (see Appendix 9 for a full list of these items). The critical HF words were identical for all three conditions. The filler item at the beginning of each block was always an HF word; either English or German, depending on the starting language. These items were additional to Sets A and B, as were the items in the practice block.

As in Experiments 1a and 1b, words were orthographically legal in either language, though some perhaps less usual (due to constraints by frequency bands). Cognate and non-cognate homographs were again excluded. The constraints posed by the frequency, letter and syllable length matching made it impossible to construct these word sets entirely of nouns, thus other syntactic categories were used as well.

Table 3.1

Mean frequencies (English: Kučera & Francis; German: Ruoff; CELEX for both English [E] and German [G]), mean letter length and mean syllable length (SD in brackets in each case) for the critical HF words, as a function of Set (A / B) and language (English / German).*

Set	Language	Kučera & Francis	Ruoff	CELEX E / G	Letter Length	Syllable Length
Set A	**English**	154.00 (200.46)	-	146.54 (207.33)	4.42 (.83)	1.42 (.50)
	German	-	103.67 (125.04)	122.54 (206.15)	4.79 (.83)	1.54 (.51)
Set B	**English**	177.33 (317.72)	-	169.88 (362.15)	4.63 (.97)	1.42 (.50)
	German	-	65.17 (88.63)	63.33 (66.34)	4.58 (.72)	1.42 (.50)

* CELEX values became available after stimulus assembly was complete; they are here included for completeness' sake only.

Note: the CELEX values for German words in Sets A and B are not significantly different: t(46) = 1.34, $p = 0.19$. The Ruoff values for the German words in Sets A and B are likewise not significantly different: t(46) = 1.23, $p = 0.23$.

Table 3.2 (below) displays the mean parameter values of the full word set per condition.

Table 3.2

Mean frequencies (English: Kučera & Francis; German: Ruoff; CELEX for both English [E] and German [G]), mean letter length and mean syllable length (all SD in brackets) for all word stimuli, as a function of Set (A / B), language (English / German) and word frequency band (HF / LF), for each condition separately (100/0; 50/50; 75/25).*

Set	Language	Frequency	Kučera & Francis	Ruoff	CELEX E / G	Letter Length	Syllable Length
				100/0			
Set A	English	HF	153.44 (159.62)	-	128.29 (147.69)	4.49 (.96)	1.31 (.47)
	German	HF	-	97.95 (171.51)	115.16 (169.09)	4.83 (1.03)	1.53 (.50)
Set B	English	HF	166.42 (173.31)	-	132.56 (194.13)	4.55 (.99)	1.34 (.50)
	German	HF	-	75.59 (107.86)	82.58 (114.81)	5.00 (.97)	1.60 (.51)
				50/50			
Set A	English	HF	141.19 (158.33)	-	122.73 (162.36)	4.52 (.85)	1.46 (.50)
		LF	3.21 (1.57)	-	3.88 (3.08)	4.52 (.77)	1.44 (.50)
	German	HF	-	121.48 (176.16)	139.17 (206.35)	4.60 (.94)	1.50 (051)
		LF	-	1.11 (.37)	6.50 (9.34)	4.60 (.71)	1.54 (.50)
Set B	English	HF	161.79 (240.18)	-	135.33 (263.45)	4.50 (.90)	1.46 (.54)
		LF	3.51 (1.99)	-	5.27 (6.15)	4.48 (.71)	1.40 (.49)
	German	HF	-	91.56 (123.84)	84.02 (130.51)	4.56 (.90)	1.54 (.50)
		LF	-	1.13 (.40)	16.34 (43.58)	4.56 (.71)	1.52 (.50)
				75/25			
Set A	English	HF	148.24 (131.80)	-	118.29 (137.73)	4.47 (.95)	1.38 (.49)
		LF	2.29 (1.46)	-	3.54 (3.08)	4.63 (.88)	1.63 (.49)
	German	HF	-	108.84 (180.52)	132.75 (189.94)	4.71 (.91)	1.54 (.50)
		LF	-	1.00 (.00)	5.21 (4.85)	4.58 (.65)	1.67 (.48)
Set B	English	HF	162.38 (196.94)	-	129.58 (219.86)	4.56 (.96)	1.40 (.52)
		LF	2.17 (.98)	-	4.29 (6.08)	4.46 (.78)	1.50 (.51)
	German	HF	-	79.44 (115.90)	77.60 (112.82)	4.76 (.96)	1.54 (.50)
		LF	-	1.00 (.00)	10.42 (12.00)	4.38 (.65)	1.46 (.51)

* Note: CELEX frequencies were not available at the time of stimulus assembly – these values are given here for comparative purposes only.

b) Nonwords:

Two sets of nonwords (sets A and B; N = 192 in each) were established, using the same method as for Experiments 1a and 1b (see Method section of Experiment 1a for details). Some of the items had been used previously, others were new, but were made up in the same manner.

In each set of nonwords there were 96 "English" nonwords (48 non-unique, 48 unique) and 96 "German" nonwords (48 non-unique, 48 unique), making a total of 192 nonwords per set. See Table 3.3 for descriptive details. All conditions saw the same nonwords, for which reason this table does not differentiate between them.

Table 3.3

Mean frequencies of words underlying 'English nonwords' (Kučera & Francis) and 'German nonwords' (Ruoff), mean position of vowel exchange (given as letter position), number of items per illegal cluster position (given as general position in the word: word initial, medial, final)[1], mean letter length and mean syllable length, as a function of stimulus set (A; B), language (English; German) and nonword type (non-unique[2]; unique[3]). Standard deviations are given in brackets.

Set	Language	NW Type	Kučera & Francis	Ruoff	Position of Vowel Exchange	Position of Illegal Cluster	Letter Length	Syllable Length
Set A	**English**	**Non-unique**	77.37 (164.16)	-	2.38 (.70)	-	4.56 (.77)	1.50 (.55)
		Unique	50.35 (86.73)	-	2.38 (.73)	I: 17 M: 15 F: 16	5.00 (.74)	1.35 (.48)
	German	**Non-unique**	-	43.93 (133.94)	2.29 (.68)	-	4.48 (.74)	1.50 (.55)
		Unique	-	20.31 (36.96)	2.67 (1.39)	I: 17 M: 17 F: 14	5.15 (.65)	1.48 (.50)
Set B	**English**	**Non-unique**	56.41 (68.24)	-	2.29 (.74)	-	4.54 (.71)	1.42 (.50)
		Unique	50.92 (99.64)	-	2.38 (.67)	I: 16 M: 15 F: 17	5.00 (.62)	1.38 (.49)
	German	**Non-unique**	-	39.37 (17.32)	2.35 (.79)	-	4.56 (.71)	1.54 (.62)
		Unique	-	24.09 (70.10)	2.42 (1.09)	I: 15 M: 18 F: 15	5.13 (.70)	1.46 (.50)

[1] since clusters are comprised of more than one letter, actual letter positions for the illegal cluster cannot be given – instead the position within the frame of the word is reported.

[2] orthographically legal in either language

[3] orthographically legal only in the target language

3.2.4 Apparatus & Procedure

The experiment was run on an IBM-compatible PC. Each participant was tested individually in a small cubicle. Stimuli were presented (in white upper-case letters) in one of the four quadrants of the large square located centrally on the black screen. Since German nouns are capitalised, all words were written entirely in capitals so as to obviate language cues (important for the language-general task). The large square measured 13cm on each side, the quadrants measured 6.5cm on each of their four sides. Responses were registered by a two-button response box, the buttons of which were labelled "+" and "-" (for "word" and "nonword" respectively) in order to be language neutral. Participants pressed "+" with the index finger and "-" with the middle finger of their dominant hand. Individuals initiated a block of experimental trials by pressing the spacebar of an adjacent computer keyboard using their non-dominant hand. 'Reaction time' was the stimulus-response interval. Stimuli remained on the screen until a response was given.

For all tasks and conditions the procedure in the experiment was as follows. The experimental programme began with written instructions in English (verbal instructions were in German, so as to activate both languages). Only participants in the language-specific task were told clearly which quadrants would have which language assigned to it. In their instructions there was a presentation of the squares with the words 'English' 'English' 'Deutsch' 'Deutsch' shown in the 'clocks' sequence (just as the stimuli would be in the experiment proper), so as to create an affiliation between the squares and their assigned languages. The instructions were followed by a block of practice trials (16 items).

Participants were informed that on each trial a letter string would appear on the screen, for which they were to decide as quickly and accurately as possible whether or not it constituted a word. Participants were not told of the proportions of HF to LF words. They were informed that words were not necessarily all nouns, but could be verbs, adverbs or adjectives as well. Participants' instructions reflected the task they were about to carry out:

Instructions for the language-specific LDT:

Participants were told to decide whether a string of letters was a real word in a specific language; two of the squares would be their "German squares" and two their "English squares". It was ensured that they had no doubt as to which language was assigned to which squares.

Instructions for the language-general LDT:

Participants were told to decide simply whether the item seen constituted a word, irrespective of its language. Participants were told not quite truthfully, that English words, German words and nonwords could appear in any of the four squares. (This deemed particularly important for those who had previously carried out the language-specific task, so that language-switches would not become obvious, in which the regularity of language-switches could have made the task less 'language-general'.)

Upon completion of the practice block, participants initiated the first of six experimental blocks by pressing the keyboard's spacebar. This triggered the first stimulus (located in the top left quadrant for all participants). For the participants in the language-specific LDT this square could be either a German or an English one (see above), so that they had to ask themselves accordingly "real German word or nonword?" or "real English word or nonword?", depending on their starting language. For the participants in the language-general LDT, the task was to decide simply "word or nonword"? Recall that the first trial of an experimental block was always a filler trial and excluded from the analysis. Upon completion of an experimental block participants were encouraged to pause before commencing with the next block. After completing the experiment (which took on average 30 minutes), each participant carried out both the English and German LLEX tests (Meara, 1994) and filled in the language background questionnaire (this took approximately a further 60 minutes).

3.3 Results & Discussion

3.3.1 Data treatment

Erroneous responses (4.7% of the data) were removed before the RT analysis. For the 100/0, the 50/50 and the 75/25 condition separately, the respective total error rates were 3.1%, 6.2% and 4.7%. On the critical HF words the error rate was very low (over all groups 1.9%), and did not vary much between the three conditions: 2.2% (100/0), 1.6% (50/50), 1.9% (75/25). Outliers (+/- 2 *SD* from each participant's cell mean) were removed (for the means analyses only) - this accounts for 3.4% of the entire data set (and for 2.8% of the critical HF set). For the 100/0, the 50/50 and the 75/25 condition the respective overall outlier rates were 3.4% [critical HF: 2.8%], 3.3% [critical HF: 3.0%] and 3.4% [critical HF: 2.8%]. Analyses were carried out both on means and medians of correct RTs. The analyses performed on median

RTs followed the same pattern as those performed on the means. To facilitate comparison between the present experiment and Experiments 1a /1b, the following reports the means analyses. Analyses were also performed on (arcsine transformed) error proportions. Inspection of the RT and error data revealed generally no speed/accuracy trade-off. Mean values presented in the text and tables are mean of means (from the subject data). Switch-costs were again calculated by subtracting the data (RT / error) of non-switch trials from that of switch trials.

The analyses comprise five sections; all are on RTs and errors. Section *3.3.2* provides information on the 'overall' analyses. This is presented first in order to provide an overview of the data. Section *3.3.3* presents the first of the two planned comparisons, contrasting the 100/0 condition against the mean of the 75/25 and the 50/50 conditions, assessing the effect presenting LF words has on the task order effect. Section *3.3.4* presents the second planned comparison, contrasting the 75/25 against the 50/50 condition, assessing the effect the proportion of LF to HF words has on the task order effect. The analyses in sections *3.3.2*, *3.3.3* and *3.3.4* are on the critical HF data only. Section *3.3.5* deals (briefly) with frequency effects, while section *3.3.6* investigates nonwords.

*3.3.2 **100/0** vs. **75/25** vs. **50/50: 'Overall' RT and error analyses on the critical HF data set***

A mixed-factor ANOVA was carried out on the set of critical HF words ('critical set' from now on). For the analyses by subjects, the within-subjects factors were Task (language-specific/language-general), Trial Type (switch/non-switch) and Language (English/German), the between-subjects factors were Condition (100/0; 75/25; 50/50) and Task Order (first task / second task). For the analyses by items, Language was a between-subjects factor, all others were within-subjects factors.

The relevant data are presented in Table 3.4. Since neither the most relevant effect (trial type), nor the factors task or language, were modified by condition, Table 3.4 does not distinguish between the three conditions.

Table 3.4

Mean correct RT (ms), SD (in brackets) and error % per cell for the critical HF words, as a function of task (language-specific/language-general), task order (first task/second task), language (English/German), and trial type (switch/non-switch).

			Trial Type				
			Switch		**Non-Switch**		
Task	**Task Order**	**Language**	**Mean RT (*SD*)**	**% Error**	**Mean RT (*SD*)**	**% Error**	**Switch-cost (RT)**
‡L-specific	**'first'**	**English**	786 (201)	2.6%	706 (168)	0.5%	80ms
		German	734 (140)	3.1%	695 (168)	1.9 %	39ms
	'second'	**English**	711 (139)	4.5%	646 (124)	1.0%	65ms
		German	670 (142)	2.8%	613 (120)	1.6%	57ms
L-general	**'first'**	**English**	689 (125)	2.1%	652 (126)	1.4%	37ms
		German	630 (938	1.7%	646 (108)	1.9%	-16ms
	'second'	**English**	685 (140)	1.6%	645 (125)	0.7%	40ms
		German	642 (109)	0.7%	638 (122)	2.4%	4ms

‡ L = language

As one may see, RTs were generally slower on switch than on non-switch trials, and this was more pronounced in the language-specific than in the language-general task. Switching seemed to be more costly on English than on German switch-trials (an 'asymmetric switch-cost'). In the language-specific, but not the language-general task, RTs were clearly slower in 'first' than in 'second' tasks. The following section provides the analytical details (note that information not directly germane to the questions posed here is presented in Appendix 5, not in the body of the following text). As shall be seen, effects visible in RT analyses are not necessarily reflected in the error analyses, suggesting that individuals placed themselves towards the cautious end of the speed/accuracy continuum.

Overall, there was no clear effect of condition [RT: $F_1(2,90) = 1.04$, $p > 0.3$; $F_2(2,188) = 16.42$, $MSE = 13021.98$, $p < 0.001$; min $F' < 1$. Errors: $F_1 < 1$; by items $p > 0.2$; min $F' < 1$]. Overall, RTs were not faster for the 'pure-HF' condition than for the 'mixed' conditions (mean RTs on critical HF, over both tasks, were: 100/0 - 689ms [135]; 75/25 - 676ms [171]; 50/50 - 658ms [115]).

It is conceivable that variability associated with English items would cloud a frequency blocking effect, such that it would be visible for the L1 only. To assess this, an analysis looking at German alone and English alone (in each case with the factors task, trial type, condition and task order) were undertaken. The analysis on the German data revealed no reliable effect of condition [$p = 0.18$ or $F < 1$]. Mean RTs, for the 100/0, 75/25 and 50/50 conditions were, respectively: 679ms (125), 657ms (161) and 640ms (102). In the analysis on the English data there was no reliable effect of condition either [$p > 0.3$ or $F < 1$]. Mean RTs to English HF words, for the 100/0, 75/25 and 50/50 conditions respectively, were: 698ms (144), 695ms (179) and 676ms (125).

This pattern is not in accord with a frequency blocking effect: faster RTs were found in the mixed conditions, not the pure condition. Individual variation may have given rise to this data pattern. Most importantly for the current purposes, these data suggest that individuals in the 100/0 condition did not base their responses exclusively on 'wordness' information (as unilinguals in Gordon's 1983 experiment were suggested to have done). However, since this is a between-subjects comparison, it does not appear sensible to read too much into this. Note, in the following, interactions between condition and any of the other experimental factors will only be mentioned if they are statistically significant.

Replicating the comparison of Experiments 1a and 1b, there was a main effect of task. Overall, RTs were slower in the language-specific task (695ms; *SD* = 159) than in the language-general task (653ms; *SD* = 129) [$F_1(1,90) = 4.42$, MSE = 75431.61, $p < 0.05$; $F_2(1,94) = 120.98$, MSE = 7361.30, $p < 0.001$; min $F'(1,97) = 4.26$, $p < 0.05$]. Participants also made somewhat more errors in the language-specific task (2.2%) than in the language-general task (1.8%), but statistically this difference was not reliably significant [$F_1(1,90) = 2.09$, MSE = 0.02, $p = 0.15$; $F_2(1,94) = 2.83$, MSE = 0.02, $p = 0.096$; min $F'(1,173) = 1.20$, $p > 0.2$].

There was, overall, a cost for switching languages. RTs were slower on switch trials (693ms; *SD* = 146) than on non-switch trials (655ms; *SD* = 136) [$F_1(1,90) = 106.34$, MSE = 2682.98, $p < 0.001$; $F_2(1,94) = 168.69$, MSE = 5922.84, $p < 0.001$; min $F'(1,173) = 65.22$, $p < 0.001$]. There were also more errors on switch trials (2.5%) than on non-switch trials (1.4%) [$F_1(1,90) = 12.25$, MSE = 0.02, $p = 0.001$; $F_2(1,94) = 15.36$, MSE = 0.02, $p < 0.001$; min $F'(1,181) = 6.81$, $p < 0.01$].

Replicating the Experiment 1a/1b comparison, there was a significant two-way interaction between task and trial type. Overall, there was a switch-cost of 60ms in the language-specific task [mean RT, switch: 725ms (162), non-switch: 665ms (151)]. In the language-general task the overall switch-cost was 17ms [mean RT, switch: 662ms (121), non-switch: 645ms (120)]. This interaction was significant [$F_1(1,90) = 46.36$, MSE = 2017.08 $p <$ 0.001; $F_2(1,94) = 46.45$, MSE = 7736.41, $p < 0.001$; min F'(1,184) = 23.20, $p < 0.001$]. In terms of errors too, there was a trial type by task interaction. Overall, there were 2.2% more errors on switch trials than on non-switch trials in the language-specific task (switch: 3.3%; non-switch: 1.1%). In the language-general task, in contrast, there were only 0.1% more errors on switch trials than on non-switch trials (switch: 1.7%, non-switch: 1.6%) [task x trial type on errors: $F_1(1,90) = 15.68$, MSE = 0.01, $p < 0.001$; $F_2(1,94) = 11.82$, MSE = 0.02, $p =$ 0.001; min F'(1,181) = 6.74, $p < 0.01$].

Task order was significant. RTs were slower in first tasks (692ms; *SD* = 152) than in second tasks (656ms; *SD* = 130) [$F_1(1,90) = 4.23$, MSE = 58828.06, $p < 0.05$; $F_2(1,94) =$ 49.11, MSE = 7890.46, $p < 0.001$; min F'(1,105) = 3.89, $p = 0.053$], but there was virtually no difference in terms of errors between first (1.9%) and second tasks (2.0%) [all Fs < 1]. This suggests that, overall, participants can grow accustomed to tasks (making the word/nonword decisions, pressing the buttons) and are thus able to react more quickly (without automatic detrimental effects to their accuracy)[49].

There was an indication of an interaction between task and task order. Overall, average RTs in language-specific tasks carried out second were 70ms faster than average RTs in language-specific tasks that were carried out first; there was virtually no difference between first and second language-general tasks (average RTs in second ones were 2ms faster). This interaction was significant by items, by subjects and min F' it was marginal [$F_1(1,90) = 2.97$, MSE = 75431.61, $p = 0.088$; $F_2(1,94) = 131.64$, MSE = 4053.62, $p < 0.001$; min F'(1,94) = 2.90, $p = 0.093$]. For errors there was no task by task order interaction [$F_1 < 1$; by items $p > 0.2$; min F'< 1].

The external account predicted an order effect (reduced language-switch cost) in the language-specific task, but only for the pure-HF condition. However, this effect was not statistically significant in this overall analysis [condition x task x task order x language-

---

[49] Incidentally, this goes some way to understanding why participants taking part in the studies carried out by, for example, Dijkstra and colleagues in the Netherlands, show faster RTs than are found for participants taking part in the experiments presented in the context of this thesis: The participants in the Dutch studies are generally recruited from the Nijmegen Department of Psychology and frequently have previous experience with LDTs (Schulpen, personal communication, March 2000). Individuals taking part in the experiments presented here were from a large variety of academic backgrounds, and none had previous LDT experience (apart from the "old" participants in Experiment 1b).

switching: $F_1(2,90) = 2.04$, MSE = 2017.08, $p = 0.1$; $F_2(2,188) = 1.42$, $p > 0.2$; min F'< 1]. The relevant data are in Table 3.5 below. In the error analyses there were no four-way interactions [$p > 0.2$ and F < 1]. This effect is investigated further in section 3.3.3, where the factor Condition does not simply contrast all three conditions, but contrasts 'pure' with 'mixed', in line with the prediction.

Additional, but not directly relevant, data are presented in Appendix 5. Other than the data presented there, there were no further significant effects / interactions in this overall analysis. The following presents the investigation into the order effects in the language-specific task, which the external account predicted.

3.3.3 100/0 vs. (75/25+50/50); planned comparison 1: Does the presence of LF words modify the task order effect?

For the analyses presented in this section, the factors of the ANOVA were Condition (100/0 vs. the mean of 75/25 and 50/50), Task Order (first / second task), Language (English / German) and Trial Type (language-switch / language-non-switch), assessing the prediction that there would be a task order effect within the language-specific task, but only for the 100/0 condition (between- and within-subject status are as for the 'overall' analysis).

The main effect of condition was significant by items, but for subjects and min F' it was not [$F_1(1,92) = 1.37$, $p > 0.2$; $F_2(1,94) = 18.95$, MSE = 15029.56 $p < 0.001$; min F'(1,105) = 1.28, $p > 0.2$]. The findings are in line with the previous discussion that there does not seem to be a frequency blocking effect here - average RTs were slower in the 'pure' (100/0) condition (689ms, *SD* = 135) than in the 'mixed' (mean of 75/25 and 50/50) conditions (668ms, *SD* = 135).

Trial type clearly affected RT (overall switch-cost: 38ms; mean RT, switch trials: 693ms, *SD* = 146; non-switch trials: 655ms, *SD* = 136) [$F_1(1,92) = 87.43$, MSE = 2707.11, $p < 0.001$; $F_2(1,94) = 124.40$, MSE = 5284.77, $p < 0.001$; min F'(1,180) = 51.34, $p < 0.001$]. In terms of errors too, there was a cost for switching languages (error percentage, switch trials: 2.5%; non-switch trials: 1.4%) [$F_1(1,92) = 11.77$, MSE = 0.02, $p < 0.001$; $F_2(1,94) = 9.16$, MSE = 0.02, $p < 0.01$; min F'(1,184) = 5.15, $p < 0.01$].

The main effect of task was significant, with overall slower RTs in language-specific than in language-general tasks (respective values: average RT = 695ms, *SD* = 159; average RT = 653ms, *SD* = 129) [$F_1(1,92) = 4.53$, MSE = 74066.70, $p < 0.05$; $F_2(1,94) = 99.58$, MSE

= 6836.49, $p < 0.001$; min F'(1,100) = 4.33, $p < 0.05$]. There were also somewhat more errors in the language-specific than in the language-general task (2.2% vs. 1.8%), but this was significant only by items, [$F_1(1,92) = 2.99$, MSE = 0.02, $p = 0.087$; $F_2(1,94) = 5.26$, MSE = 0.02, $p < 0.025$; min F'(1,170) = 1.91, $p > 0.1$].

The main effect of task order was (marginally) significant, with slower RTs (by 36ms) in first tasks (average RT: 692ms, *SD* = 152) than in second tasks (average RT: 656ms, *SD* = 130) [$F_1(1,92) = 3.70$, MSE = 58261.42, $p = 0.057$; $F_2(1,94) = 46.02$, MSE = 5933.47, $p < 0.001$; min F'(1,107) = 3.42, $p = 0.067$]. On its own, such a pattern could be regarded simply as a practice effect. However, it is the interaction of task order with the other factors that is of interest here (see below). In terms of errors, task order was not significant [Fs < 1] - 1.9% errors in first tasks and 2.0% errors in second tasks.

Two of the two-way interactions are of interest. Firstly, task x trial type. As found for Experiments 1a vs. 1b, the cost for switching languages was significantly greater in the language-specific task (60ms) than in the language-general task (17ms) - language-specific, switch: 725ms (162), non-switch: 665ms (151); language-general, switch: 662ms (121), non-switch: 645ms (120) [$F_1(1,92) = 42.78$, MSE = 1986.34, $p < 0.001$; $F_2(1,94) = 50.09$, MSE = 5489.25, $p < 0.001$; min F'(1,185) = 23.07, $p < 0.001$]. For errors there was also a greater switch-cost in the language-specific task (2.2%) than in the language-general task (0.1%) - language-specific, switch: 3.3%, non-switch: 1.1%; language-general, switch: 1.7%, non-switch: 1.6% [$F_1(1,92) = 18.03$, MSE = 0.01, $p < 0.001$; $F_2(1,94) = 9.17$, MSE = 0.02, $p < 0.01$; min F'(1,169) = 6.08, $p < 0.001$].

Secondly, task x task order. 'Language-specific second' tasks produced faster RTs than 'language-specific first' tasks (an average of 70ms faster). For language-general tasks there was virtually no difference between first and second tasks (the latter were on average 2ms faster) [task x task order: $F_1(1,92) = 1.82$, $p > 0.1$; $F_2(1,94) = 76.05$, MSE = 3465.76, $p < 0.001$; min F'(1,96) = 1.78, $p > 0.1$]. This pattern had shown a greater trend towards significance in the overall analysis, suggesting that larger group sizes might be needed to achieve greater power. Nonetheless, this trend is of interest, as it indicates that the task order effect (above) was not simply a practice effect. For errors there was no significant task by task order interaction [by subjects and items: $p > 0.1$; min F' < 1].

There were no other significant two-way interactions involving the factors condition, task, task order and trial type. The next relevant interaction concerns the prediction made at the outset:

This prediction was that the cost for switching languages would be smaller in a 'language-specific second' task compared to a 'language-specific first' task, but only when the stimulus list excluded LF words (i.e., only for the 100/0 condition). This supposition is addressed by a four-way interaction between the factors condition, task, task order and trial type. This interaction was marginally significant by subjects, by items and min F' it was only a trend / n.s. [$F_1(1,92) = 3.77$, MSE = 1986.34, $p = 0.055$; $F_2(1,94) = 1.73$, MSE = 8635.80, $p > 0.1$; min $F'(1,168) = 1.19$, $p > 0.2$]. Given that there were 16 participants in each task order cell, there may not have been enough power for the trend to have become more established. (See Table 3.5 for the relevant data. Note, the two mixed conditions are displayed separately there, so that this table can also be used as reference for the second planned comparison, which contrasts the 75/25 and the 50/50 conditions.) In terms of errors there was no four-way interaction [all Fs < 1].

Is the difference between the conditions indicated by the above (RT) four-way interaction that, indeed, only the 'pure' condition showed a task order effect (and only in the language-specific task) ?

For the 100/0 condition, the three-way interaction between task, task order and trial type was significant by subjects, by items it was at least marginal [$F_1(1,30) = 5.29$, MSE = 1574.40, $p < 0.05$; $F_2(1,94) = 3.28$, MSE = 6182.12 $p = 0.074$; min $F'(1,111) = 2.02$, $p > 0.1$]. For this condition, in the language-specific task, the switch-cost was larger (by 31ms) when this task came first than when it came second (first, switch: 782ms, *SD* = 126, non-switch: 694ms, *SD* = 102; second, switch: 724ms, *SD* = 112, non-switch: 667ms, *SD* = 95). In contrast, in the language-general task the switch-cost was slightly smaller (by 9ms) when this task came first than when it came second (first, switch: 676ms, *SD* = 94, non-switch: 672ms, *SD* = 106; second, switch: 665ms, *SD* = 100, non-switch: 652ms, *SD* = 104). In terms of errors there was no significant effect [all Fs < 1].

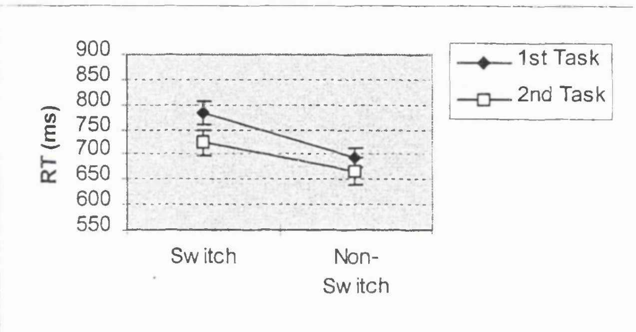
Critically, for the mixed conditions (taking the mean of the 75/25 and 50/50 conditions), the three-way interaction between task, task order and trial type was not significant [Fs < 1 or *p*s > 0.2, for RTs and errors]. For the mixed conditions, in the language-specific task, the switch-cost was very similar (11ms smaller) when this task came first compared to when it came second (first, switch: 742ms, *SD* = 76, non-switch: 684ms, *SD* = 68; second, switch: 684ms, *SD* = 77, non-switch: 615ms, *SD* = 72). In the language-general task switch-costs were also very similar (4ms smaller) when this task came first than when it came second (first, switch: 650ms, *SD* = 66, non-switch: 631ms, *SD* = 57; second, switch: 666ms, *SD* = 63, non-switch: 643ms, *SD* = 62).

Figures 3.1 *a)* and *b)* show the switch-cost patterns in the 'first' vs. the 'second' language-specific tasks, for the 'pure' and 'mixed' conditions separately. (See Table 3.5 for the full data patterns.) Figures 3.2 *a)* and *b)* show the switch-cost patterns in the 'first' and 'second' language-general tasks, for the 'pure' and 'mixed' conditions separately.

Figure 3.1

Mean correct RT (ms) of the critical HF words, in the language-specific task, on switch and non-switch trials, as a function of task order, for 'pure' and 'mixed' conditions separately. Error bars are standard error of the mean.

a) the 100/0 condition



b) the mean of the 75/25 and 50/50 conditions

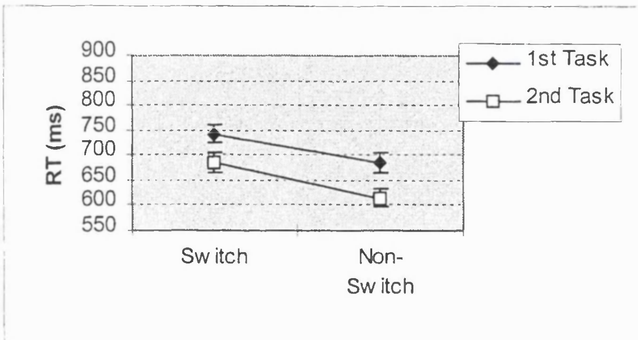


Table 3.5

Mean correct RT (ms), standard deviations (in brackets) and error percentages per cell for the critical HF words, as a function of task (language-specific/language-general), task order (first task/second task) and trial type (switch/non-switch), for each condition separately (the 75/25 condition is reported here for completeness' sake).

100/0:

		Trial Type				
Task	**Task Order**	**Switch**		**Non-Switch**		**Switch-Cost**
		Mean RT (*SD*)	**% Errors**	**Mean RT (*SD*)**	**% Errors**	significance by subjects / items
Language-specific	**First Task**	774 (130)	3.4	699 (118)	0.8	75ms **** / **** 2.6% † / *
	Second Task	711 (144)	5.5	674 (148)	2.3	37ms *** / **** 3.2% *** / ***
Language-general	**First Task**	678 (136)	2.1	671 (147)	2.6	7ms †† / †† -0.5% †† / ††
	Second Task	657 (102)	1.0	643 (115)	1.0	14m † / † 0.0% †† / ††

50/50:

		Trial Type				
Task	**Task Order**	**Switch**		**Non-Switch**		**Switch-Cost**
		Mean RT (*SD*)	**% Errors**	**Mean RT (*SD*)**	**% Errors**	significance by subjects / items
Language-specific	**First Task**	737 (152)	2.1	693 (134)	1.6	44ms ** / *** 0.5% †† / ††
	Second Task	657 (110)	2.9	593 (72)	1.3	64ms *** / **** 1.6% † / †
Language-general	**First Task**	658 (100)	1.8	645 (101)	0.5	13ms † / † 1.3% † / †
	Second Task	648 (93)	1.0	632 (94)	1.8	16ms † / †† -0.8% † / †

75/25:

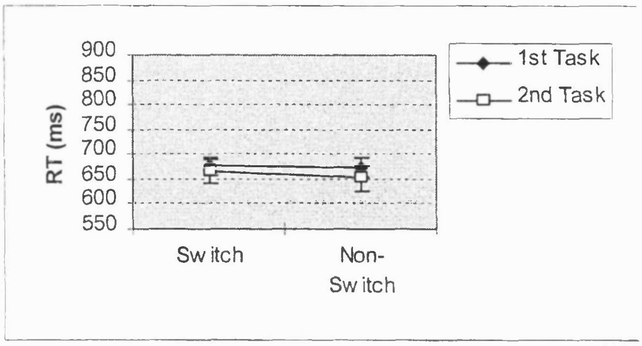
		Trial Type				
Task	**Task Order**	**Switch**		**Non-Switch**		**Switch-Cost**
		Mean RT (*SD*)	**% Errors**	**Mean RT (*SD*)**	**% Errors**	significance by subjects / items
Language-specific	**First Task**	769 (229)	3.1	708 (232)	1.3	61ms **** / **** 1.8% * / †
	Second Task	704 (163)	2.6	621 (124)	0.3	83ms **** / **** 2.3% * / **
Language-general	**First Task**	643 (108)	1.8	631 (94)	1.8	12ms † / † 0.0% †† / ††
	Second Task	685 (171)	2.6	648 (156)	1.8	37ms **** / *** 0.8% †† / ††

**** $p < 0.001$
*** $p < 0.01$
** $p < 0.025$
* $p < 0.05$
† $p > 0.05$
†† $F < 1$

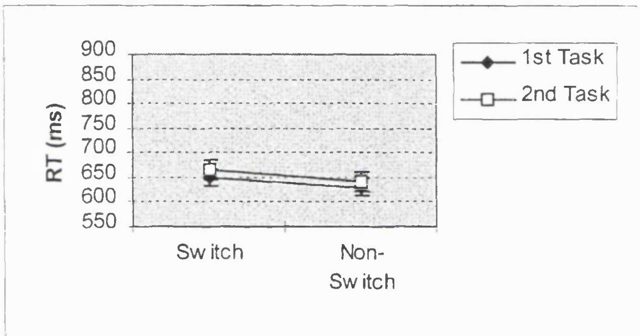
Figure 3.2

Mean correct RT (ms) of the critical HF words, in the language-general task, on switch and non-switch trials, as a function of task order, for 'pure' and 'mixed' conditions separately. Error bars are standard error of the mean.

a) the 100/0 condition



b) the mean of the 75/25 and 50/50 conditions



In sum, given a 'pure-HF' stimulus list, the language-switch cost in the language-specific task appeared to be affected by a preceding language-general task, so as to reduce language-switch costs; given a 'mixed' stimulus list, there was no evidence of such an effect. The external account had predicted this type of pattern, saying that a relaxed response criterion would permit placing more weight on wordness than on language signals. The internal account had predicted the different conditions to perform similarly (as had the input-switch account).

In various instances the expected pattern is statistically significant only in the subject, but not in the item (or the min F') analyses. It is possible that more items (in each language/trial type cell, for each task order and each frequency condition) are required to establish the effect more firmly. Furthermore, it is possible that the results are not robust because even in the condition that was expected to show the task order effect (i.e., pure-HF condition), due to word-like nonwords, participants may have placed sufficient weight on language signals to dilute the difference between them and the mixed-frequency condition. Another possibility is that individuals' L2 language abilities differed (with perhaps more balanced bilinguals in the mixed conditions than in the pure condition). The language background information does not indicate that this is the case, but it may be that the language tests employed were not sufficiently sensitive (this is further discussed in Chapter 6, at the end of section *6.2.4.1*).

3.3.4 75/25 vs. 50/50; planned comparison 2: Do frequency proportions modify the task order effect?

Regarding this subsidiary question, whether the actual proportions of HF to LF words would give rise to 'proportion effects' (or whether there was just a pure/mixed difference), the two 'mixed' conditions were contrasted with one another in a mixed-factor ANOVA with the factors Task (language-specific / language-general), Trial Type (language-switch / language-non-switch), Language (English / German) Task Order (first task / second task), and Condition (50/50; 75/25); between- and within-subject status are as described in section 3.3.2

.

The main effect of condition was not significant by subjects or min F', though it was significant by items [$F_1 < 1$ $F_2(1,94) = 7.95$, MSE = 6004.55, $p < 0.01$; min F' < 1]. Average RTs were 676ms (171) in the 75/25 condition and 658ms (115) in the 50/50 condition. In terms of errors there was statistically no difference between the two mixed conditions [$p > 0.2$ and $F < 1$] - 75/25: 1.9% errors; 50/50: 1.6% errors.

The main effect of task order was significant by items, by subjects and min F' it was only marginal [$F_1(1,60) = 3.60$, MSE = 48537.81, $p = 0.063$; $F_2(1,94) = 34.76$, MSE = 6618.11, $p < 0.001$; min F'(1,73) = 3.26, $p = 0.075$] - on average RTs were faster in first tasks (685ms, *SD* = 158) than in second tasks (648ms, *SD* = 130). In terms of errors there was no effect of task order [Fs < 1].

The main effect of trial type was significant, with faster RTs on non-switch (656ms, *SD* = 122) than on switch trials (690ms, *SD* = 129) [$F_1(1,60) = 79.86$, MSE = 2716.41, $p < 0.001$; $F_2(1,94) = 85.20$, MSE = 8028.36, $p < 0.001$; min F'(1,146) = 41.22, $p < 0.001$]. There were more errors on language-switch (2.2%) than on language-non-switch trials (1.3%)

[$F_1(1,60) = 7.48$, MSE = 0.02, $p < 0.01$; $F_2(1,94) = 7.34$, MSE = 0.02, $p < 0.01$; min F'(1,146) = 3.70, $p = 0.056$].

As found previously, the effect of language-switching was greater in the language-specific than in the language-general task (63ms and 19ms, respectively) [task x trial type, RT: $F_1(1,60) = 26.64$, MSE = 2238.42, $p < 0.001$; $F_2(1,94) = 20.12$, MSE = 8754.22, $p < 0.001$; min F'(1,153) = 11.46, $p < 0.001$] - mean RTs on switch and non-switch trials were 717ms (172) vs. 654ms (158) in the language-specific task, and 658ms (122) vs. 639ms (113) in the language-general task. In terms of errors there was a task by trial type effect, which was only marginal by items though [$F_1(1,60) = 5.60$, MSE = 0.01, $p < 0.025$; $F_2(1,94) = 3.46$, MSE = 0.02, $p = 0.066$; min F'(1,155) = 2.14, $p = 0.15$] - in the language-specific task average error rates were 2.7% and 1.1% on switch and non-switch trials respectively; the corresponding data in the language-general task was 1.8% and 1.5%.

Critically, the three-way interaction between task, task order and trial type was not significant, either for RTs or errors [Fs < 1 or *p*s > 0.2], suggesting, as hypothesised in the Introduction to this chapter, that the presence of LF words (in company with word-like nonwords) blocks using a relaxed response criterion. Had such a relaxed response criterion been used, participants in the 'mixed' conditions would presumably have shown the same switch-cost reduction in the 'language-specific second' task as was seen for the 'pure' condition.

None of the above effects were modified by the factor condition. Critically, the four-way interaction between condition, task order, task and trial type was n.s. [RTs: Fs < 1; errors: *p*s > 0.2]. This suggests that receiving a larger proportion of HF words (i.e., the 75/25 vs. the 50/50 condition) does not change the picture described in the preceding paragraph.

Additional analyses (pertaining to language abilities) are to be found in Appendix 5a.

In sum, where there was an indication of different task order effects for the pure and mixed conditions, there was none between the two mixed conditions. The data patterns are in accord with the external account's hypothesis that a pure-HF list allows placing more weight on wordness than on language signals, in which case language-switch costs may be reduced, but the presence of word-like nonwords may block making even further use of wordness signals. The presence of LF words, regardless of their proportion, makes a more careful response criterion necessary. Thus a picture emerges of a bilingual language processing system which is very sensitive to the characteristics of the input (bottom-up) as well as to task instructions. This could be investigated further by contrasting a pure-HF condition in which nonwords are very word-like (as here) with a pure-HF condition in which nonwords are less word-like. The hypothesis would be that when nonwords are less word-like, individuals are able to make more use of wordness signals when nonwords are more word-like. In this scenario, a greater task order effect could be expected for the former condition compared to the latter.

3.3.5 Word frequency effects (RT and error analyses) - HF vs. LF words

These data are subsidiary to the preceding word analyses. They provide background information on the participant group. For the frequency analysis a subset of HF and LF items common to the 50/50 and the 75/25 condition was established. Where degrees of freedom vary in the RT item analysis, this is due to removal of errors and outliers, which reduced some cells to zero. The design (for the analyses by subjects) was a mixed-factor ANOVA with the within-subjects factors Task (language-specific / language-general), Language (English / German), Frequency (HF / LF) and Trial Type (language-switch / language-non-switch), and the between-subjects factors Condition (50/50, 75/25) and Task Order (first task / second task). For the analyses by items, Language and Frequency were between-subject factors, all others were within-subjects.

Experiments 1a and 1b had shown an effect of frequency. HF words had faster RTs than LF words, and this difference had been greater for English than for German. An important finding had been that frequency had interacted neither with trial type nor with task. The same pattern was found here.

Overall the effect of frequency was such that RTs were faster on HF words (686ms [145]) than on LF words (841ms [246]) [$F_1(1,60) = 411.96$, MSE = 15055.92, $p < 0.001$; $F_2(1,174) = 370.36$, MSE = 50897.08, $p < 0.001$; min $F'(1,188) = 195.03$, $p < 0.001$]. There were fewer errors on HF words (1.7%) than on LF words (16.3%) [$F_1(1,60) = 512.60$, MSE = 0.03, $p < 0.001$; $F_2(1,188) = 137.80$, MSE = 0.31, $p < 0.001$; min $F'(1,247) = 108.60$, $p < 0.001$].

The effect of frequency was greater for English than for German. On average, RTs were 705ms (153) and 936ms (283) to English HF and LF respectively. For German they were 667ms (133) and 747ms (152) for HF and LF respectively [frequency x language: $F_1(1,60) = 157.43$, MSE = 9343.76, $p < 0.001$; $F_2(1,174) = 61.05$, MSE = 50897.08, $p <$

0.001; min F'(1,233) = 43.99, $p < 0.001$]. The error analysis showed the same picture. Overall, there were many more errors on English LF than HF words (27.5% versus 1.7%), while for German the difference was not very great (5.1% versus 1.9%) [frequency x language: $F_1(1,60) = 213.66$, MSE = 0.03, $p < 0.001$; $F_2(1,188) = 79.85$, MSE = 0.31, $p < 0.001$; min F'(1,247) = 58.13, $p < 0.001$]. These findings are in line with the language background information (see Method section), which had indicated overall somewhat greater knowledge/certainty of German compared to English (this knowledge asymmetry could be exaggerated on LF words).

As in Experiments 1a and 1b, there was no interaction between frequency and trial type [for both RT and errors Fs < 1 or $p > 0.2$] and none between frequency and task [for both RT and error analyses Fs < 1 or $p > 0.2$]. These findings were the same for both the 75/25 and the 50/50 conditions.

3.3.6 RT and error analyses for nonword trials

For the subject analyses, the within-subjects factors of this mixed-factor ANOVA were Task (language-specific/language-general), Language (English/German), Trial Type (switch/non-switch) and Nonword Type (unique/non-unique), the between-subjects factors were Condition (100/0; 50/50; 75/25) and Task Order (first task / second task). For the item analyses, Language and Nonword Type were between-subjects factors, all others were within-subjects. Where degrees of freedom vary in the item analyses, this is due to missing data points which reduced some cells to zero.

Note that when the term 'English non-unique NW' is used in the following, this means simply that the nonword was presented on an English trial, the stimulus itself was orthographically neutral to language (e.g., HOND). 'English unique NW' refers to a nonword that was orthographically legal only in English (e.g., SWERM; such items were only presented on 'English' trials). The corresponding applies for German. Table 3.6 presents nonwords' mean correct RTs, *SD* and error % as a function of task, language, nonword type and trial type.

Table 3.6

Mean correct RT (ms), standard deviations (in brackets) and error percentage per cell for nonwords, as a function of task (language-specific / language-general), language (English / German), nonword type (non-unique[1] / unique[2]) and trial type (switch / non-switch). Switch-costs have been included to facilitate reading the table.

			Trial Type						
			Switch			Non-Switch			
Task	**Language**	**Nonword Type**	**Mean RT**	**(*SD*)**	**% Error**	**Mean RT**	**(*SD*)**	**% Error**	**Switch cost (ms)**
Lang.-specific	**English**	**Non-Unique**	963	(290)	3.4	914	(304)	3.5	49
		Unique	961	(334)	4.7	893	(290)	5.7	68
	German	**Non-Unique**	923	(234)	2.9	906	(250)	2.5	17
		Unique	936	(239)	3.2	878	(222)	2.5	58
Lang.-general	**English**	**Non-Unique**	883	(229)	3.9	878	(224)	3.0	5
		Unique	879	(245)	5.9	863	(224)	5.7	16
	German	**Non-Unique**	882	(226)	5.6	921	(291)	5.6	-39
		Unique	899	(224)	3.5	858	(201)	2.6	41

[1] 'non-unique': orthographically legal in either language; [2] 'unique': orthographically legal only in the target language

Note: since nonword type effects were not modified by condition, Table 3.6 does not differentiate between them, but instead presents the average values of all conditions per cell. Nonword type did not interact with task either, but as each task is represented in a separate table in Chapter 2, the same is done here to allow comparison.

The data pattern of Table 3.6 shows generally smaller language-switch costs in the language-specific than the language-general task. Overall, RTs for non-unique nonwords appear to be slower than those for unique nonwords, though this is not always the case (German, switch trials). As mentioned at the outset, these nonword data are subsidiary to the word data. For brevity's and clarity's sake only the relevant issues are discussed here. The first section deals with the switch-cost, the second with the effect of nonword type. Where statistical details have been abbreviated (or factors have not been reported), details are to be found in Appendix 6.

3.3.6.1 The effects of switching on nonword trials

RTs were significantly slower on switch than on non-switch trials [$p < 0.001$ for F_1, F_2 and min F'; switch: 916ms (256); non-switch: 889ms (253)], but switching did not affect accuracy much [$p = 0.054$ by subjects; by items and on min F' $F < 1$; switch: 4.2% errors, non-switch: 4.0% errors].

Most importantly, and replicating the nonwords' task comparisons in Chapter 2, there was a significant interaction (RT analyses) between task and trial type [$p < 0.001$ by subjects, items and min F'] (details in the next paragraph). In terms of errors there was no interaction between task and trial type [all Fs < 1]. So far this is in accord with both the external and the internal accounts. However, the following data distinguish between the two (see Appendix 6 for full details).

In the language-specific task, there was a significant effect of trial type [mean RTs to nonwords: 946ms (277) and 898ms (268) on switch and non-switch trials respectively; $p < 0.001$ by subjects, items and min F']. This is in accord with the notion of having to switch between (strongly) mutually inhibited task schemata (external account), or of having to overcome strong top-down inhibition (internal account). In the language-general task there was virtually no difference between switch and non-switch trials (886ms [230] and 880ms [238] respectively) [$p > 0.2$ by subjects; $p = 0.068$ by items, min F' < 1]. This finding can be accommodated by the external account, which proposed that the switch-cost to *words* in the language-general task was in part due to language signals feeding into the decision process paralexically. As argued in the Introduction to Chapter 2, *nonwords* are unlikely to have very clear language signals (on the one hand because only their real-word neighbours are represented in the lexicon, and on the other hand because half of the nonwords, all non-unique ones, will presumably give rise to language signals from both languages). Without clear language signals, there is less evidence biasing towards a given response paralexically, and so there is less response competition. The task schema component of this cost may be small, as nonwords can theoretically be correctly rejected by either task schema; in fact, when a nonword is presented on a language-switch trial, there may not even be a change of task schema (in the language-general task). As discussed in Chapter 2, the internal account predicted a switch-cost for nonwords in the language-general task, since, just as is the case for words, on switch trials, the activation decrement needs to be overcome in order for the participant to ascertain whether the letter string has a representation in the lexicon or not.

The task by trial type interaction was not modified by condition or task order [in both the RT and error analyses either Fs < 1 or $p > 0.2$].

3.3.6.2 Nonword type effects

RTs were slower for non-unique nonwords than for unique ones [909ms (258) and 896ms (252) respectively]. This difference was significant only by subjects, [$p < 0.01$, $p > 0.1$ and $p > 0.1$, for subjects, items, and min F', respectively]. There were somewhat more errors on unique than non-unique nonwords (4.2% versus 3.9%) [$p < 0.025$ for subjects and items, but for min F' $p > 0.1$]. In Experiments 1a and 1b the nonword type effect had overall not been statistically significant. Perhaps nonword data of this type are very noisy, such that the larger subject group in the present experiment (N = 96) was better able to indicate an effect which could not emerge in the preceding experiments (N = 20 per experiment). The nonword type effect could reflect either cross-language influences (in which case it would provide support for the non-selective access account), or it could reflect recognition processes, in which case it would not provide support for the non-selective access view, but would not rule it out either.

As found previously in the comparison between Experiments 1a and 1b, the nonword type effect was not modified by task [RT: all Fs < 1; errors: $p < 0.025$ by subjects, $p = 0.082$ by items, n.s. on min F']. As stated in Chapter 2, this is in accord with the external account. It is also in accord with the internal account if the nonword type effect described here is not a result of cross-language influences.

The following interactions suggest that the nonword type effect found here must be something other than cross-language influences.

Nonword type interacted with trial type [$p < 0.001$ by subjects, items and min F']. This interaction had not been significant in Experiments 1a or 1b. In the present experiment there was very little difference between unique and non-unique nonwords on switch trials (the unique ones were 6ms slower), whereas on non-switch trials unique nonwords had an advantage over non-unique ones (unique NWs were on average 32ms faster). In the error analysis there was no interaction between nonword type and trial type [all Fs < 1]. The RT findings differ from those in Experiments 1a and 1b. There, there had been no significant interaction between nonword type and trial type. Experiment 2, with more participants, may have been more sensitive than Experiments 1a and 1b. If the nonword type effect reflected cross-language influences, then one would have expected this to be present regardless of trial type. If the nonword type effect reflects recognition processes, then one may conjecture that recognition processes are more effective when an individual is 'in' a language, compared to when he has just switched into it. However, being post-hoc, these suggestions are only speculative.

The nonword type x trial type interaction was not modified by task [Fs < 1 or $p > 0.2$, on RT and errors].

The effect of nonword type was modified by task order [significant by subjects and items, marginal on min F': $p < 0.025$, $p < 0.001$ and $p = 0.059$ respectively]. In 'first' tasks RTs were virtually the same for unique and non-unique nonwords (unique ones were 2ms faster), whereas in 'second' tasks RTs to unique nonwords were clearly faster than RTs to non-unique ones (by 24ms). In the error analyses nonword type was not modified by order [$p > 0.17$ by subjects and items; min F' < 1]. The RT data suggests that in the first task individuals took unique orthography to indicate 'wordness'. It is possible that such unique orthography initially biased towards a 'yes' response paralexically. To avoid errors, the activation level of 'no' needed to exceed that of 'yes'. With time individuals appear to have learned that unique orthography was not a good indicator of wordness. They could then have reduced the weight given to unique nonwords' signals in the paralexical decision process.

The effect of nonword type was modified by language in the error analyses [all $p < 0.001$]. There were 5.5% and 3.6% errors on English unique and non-unique nonwords respectively. For German the respective data were 3.0% and 4.2%. (For each language separately, the effect of nonword type on error rates was significant.) In the RT analysis there was no interaction between language and nonword type [all Fs < 1]. The error pattern suggests that individuals used unique orthography as an indication of lexicality (i.e., wordness) for their weaker language (English). See also Thomas & Allport (2000) for a similar pattern (albeit with greater error rates). The reason for the larger error rate on German non-unique nonwords (compared to unique nonwords) is unclear.

Note, none of the effects or interactions discussed here were modified by condition.

In sum, the nonword data are in accord with the external account's task schema notion. For the internal account the matter is less clear-cut. If nonwords of this type give rise to cross-language influences (a question which it is not entirely possible to answer here), then the internal account is not in accord with part of the data (i.e., the lack of a nonword type by task interaction). If nonwords of this type do *not* give rise to cross-language effects, then the internal account can accommodate the lack of a nonword type by task interaction. What remains a problem for the internal account in any event, is that there was no evidence of a switch-cost on nonword trials in the language-general task, even though there was such a cost on word trials.

3.4 General Discussion

Experiment 2 explored the difference between the two versions of the bilingual LDT (language-specific / language-general). In the language-specific task, cued by position in the square, participants decided whether or not an item was a word in the designated language. In the language-general task, the decision had simply to be 'word or nonword', irrespective of language (there was no assignment of languages to squares in this task). The critical HF words were presented in three different frequency contexts 1) all words were HF ('pure-HF' or '100/0' condition); 2) words were half HF and half LF ('50/50' condition); 3) words were three quarters HF and one quarter LF ('75/25' condition).

The predictions made at the outset of this chapter and the general data patterns obtained will be discussed in turn in the following.

3.4.1 Word data

It had been predicted that Experiments 1a and 1b would be replicated in terms of 1) a language-switch cost being present and 2) this cost being modified by task. Both were found to be the case (both in terms of RTs and errors). In accord with the external and internal (but not the input-switch or the associative connections) accounts, the cost for switching languages was greater in the language-specific task than in the language-general task.

The 'new' effects pertained to the inclusion of the factors task order (first / second task) and condition (100/0; 50/50; 75/25). The external account predicted that there would be a smaller language-switch cost for 'language-specific second' than for 'language-specific first', but only for the 'pure-HF' condition, not for the 'mixed' conditions (50/50 and 75/25). This was based on the different weights individuals would place on wordness signals and language signals, given their stimulus list composition and their previous task experience. Individuals in the 'pure-HF' condition carrying out a language-general task as their first task, would learn to place much weight on wordness signals - and continue to do so in the following language-specific task. In contrast, individuals in a 'mixed' condition, in their language-general task, would adopt a comparatively more stringent response criterion (which entailed stronger language signals), and not place much weight on wordness signals (as these could not reliably differentiate between LF words and word-like nonwords. In the following language-specific task they would continue in this manner, both because of task constraints but also because of the item constraints already described. On this basis, the 'mixed' conditions would not be able to reduce the language-switch cost in a 'language-specific second' task compared to a 'language-specific first' task. For the 'mixed' conditions the task x task order x trial type interaction was n.s. [Fs < 1 or *p*s > 0.2]. For the 'pure-HF' condition, in contrast, there was such a difference (see Table 3.5). However, it was statistically significant only by subjects, by items it was marginal, and on min F' it was n.s. It may be that variability over items was too great, which could perhaps be rectified by presenting greater numbers of items per cell (language/trial type) - for each task order and frequency condition. Another possibility is that, due to word-like nonwords, sufficiently thorough processing took place to allow language signals to influence responses paralexically, reducing the time-saving effect of using wordness information paralexically.

3.4.1.1 What role do language signals play?

External account. Recall that the critical HF words were themselves proceeded by HF words. That means that while *language* switched on these trials, the required response was always the same as for the previous item (response-repetition). In the task where language information was important (language-specific task) and task cues were given (position cueing whether a decision was to be made in German or English), it was, apparently, more difficult to respond quickly and accurately on language-switch / response-repetition than on language-non-switch / response-repetition trials. In contrast, when the task did not require taking language information into account (language-general task), RTs and errors were only slightly greater on language-switch / response-repetition than on language-non-switch / response-repetition trials. Language signals (from the word) as well as task cues (from the position in the square) indicate that 'a change has taken place' and could contribute to reconfiguring the task schema (and/or directly feed into the units underlying a particular response, biasing towards it, presumably by raising its activation level). When language-switch and response-repetition coincided, responses were delayed. Such a pattern is in accord with the suggestion that the signal 'a change has taken place' biases towards the response previously *not* used, which in the case of a response-repetition is incorrect. To avoid an error, this bias has to be revoked (once the processing of the letter string has provided the necessary information to show what the correct response should be), resulting in a delay. There is supporting evidence for this conjecture in the literature. Thomas & Allport (2000) found, in a bilingual language-switching study requiring lexicl decisions, that, where responses were the same on adjacent trials ('yes', 'yes'), RT was faster on language-non-switch than on language-switch trials. In contrast, where the currently required response was *different* from the preceding one ('no', 'yes'), RT was *slower* on language-non-switch than on language-switch trials. This pattern was significant in their language-specific LDT, in their language-general LDT it was a trend (Thomas & Allport's 2000 Experiments 2 and 1 respectively). (See also Rogers & Monsell, 1995, for the corresponding pattern in a task-switching experiment.)

Each item's position in the square furnishes a task cue in the language-specific task (indicating whether a decision is currently required for German or for English). The task cue could, just as the language signal, help to establish the required task schema. And further, also as just described for the language signals, the paralexical decision process could be using the information deriving from the task cue to feed into the calculation of the most probable outcome - and hence bias towards one response or the other. Just as with the information derived from language, this bias would be helpful on language-switch/response-switch trials, but a hindrance on language-switch/response-repetition trials. This task cue would be available only in the language-specific, not the language-general task (in the latter case words were also presented in the 'clocks' fashion, but there was no assignment of language to position). A task cue would be available early on, feeding into the paralexical decision process and could thus potentially have a pronounced effect in the response system, that is, a language-switch cost may not just be related to language signals, but also to signals from task cues.

The small switch-cost in the language-general task may relate to task schemata. That is, if they are mutually inhibited: overcoming that (small) inhibition; if they are jointly active: raising the activation level of the 'new' schema above that of the 'old' one. Another component may come from paralexical decision processes, which serve to bias towards either a 'yes' or a 'no' response, depending on what signals are dominant (wordness / 'language-change'). If there were only one task schema, then the switch-cost could potentially be accommodated by the paralexical component alone. This could be envisaged as follows. A change of language is registered by the participant, and feeds into the response mechanism as the message "a change has occurred". Such a message could indicate that the current response should be different from the previous one[50]. In the case of two successive HF words (i.e., two consecutive 'yes' responses), activating 'no' rather than 'yes' is erroneous. To avoid mistakes, individuals need to ensure that 'yes' dominates over 'no' (which has already received some activation). Re-adjusting this activation balance would take time. That is, a switch-cost would also be conceivable when there is only one task schema.

The internal account is able to predict a smaller language-switch cost for 'language-specific second' compared to 'language-specific first'. However, it then predicts that this pattern should be found for all conditions. This was not the case.

The input-switch account predicts that if there are order effects, they should be present equally for the language-specific and the language-general task. If previous task experience helped to speed up the process of re-setting the input-switch (which could be the reason for 'language-specific second' having smaller switch-costs than 'language-specific first'), then there should also have been such an order effect for the language-general task. There was not. It also cannot predict the fact that only the 'pure-HF' condition showed this effect of task order.

In sum, the word data were best predicted by the external account. The internal account (likewise the input-switch account) was only able to capture aspects of the data patterns.

3.4.2 Nonword data

The principal findings concern the effects of language-switching and nonword type.

For nonwords, switching languages carried a cost in the language-specific, but not in the language-general task. This is in accord with the external account. The internal account predicted that there should be some cost for switching on nonword trials, just as there was for word trials, in the language-general task. In Experiment 1a (language-specific task) the effect of switching was not statistically significant. It is possible that, due to the larger participant group in Experiment 2's language-specific task, the effect which had not been statistically

---

[50] In normal bilingual language processing one may observe that a change of input language frequently triggers a change of output language (both personal experience and observation). Though in a computerised LDT the language of output cannot change, it may be that what would otherwise have been a change of language is transferred onto the button response, such that there is a change of response.

significant previously (Experiment 1a) was significant here.

There was an effect of nonword type in Experiment 2. If the nonword type effect is not solely a reflection of some recognition processes, but also carries in it a component of cross-language influences, then this finding constitutes support for a non-selective access view. This finding on its own cannot distinguish between the external and internal accounts. In accord with the external account's task schema notion, nonword type and task did not interact. The internal account can only account for this finding if the nonwords used here do not elicit cross-language effects. However, if such cross-language effects *are* part of the nonword type effect, then the internal account would predict an interaction between nonword type and task (smaller nonword type effect in the language-specific task, where suppression of the non-target language is stronger).

In sum, the nonword data provide support a) for the external account and b) possibly for a non-selective access view. They do not unequivocally support the internal account.

3.4.3 Differences between Experiments 1a/1b and 2

In Experiment 2 the language-switch costs over all three conditions were 60ms (language-specific) and 17ms (language-general). These costs are smaller than the switch-costs in Experiments 1a and 1b (117ms and 38ms respectively on the full data set; 192ms and 34ms respectively for the HF subset). To some extent the difference between Experiments 1a /1b and 2 may be related to individual differences. However, procedural differences may have contributed to this seeming reduction in cost. In Experiment 2 there were 'general changes' on each trial (a new position; see the discussion at the end of Chapter 2). In Experiment 1a there was a 'general change' only on language-switch trials (colour change). Thus it may be that there was essentially a greater difference between switch and non-switch trials in Experiment 1a than in Experiment 2's language-specific task, with a correspondingly greater language-switch cost[51]. Another difference between Experiments 1a/1b and 2 is that in the former language-switch costs were symmetric, while they were asymmetric in Experiment 2. A plausible explanation (discussed in more detail in Chapter 6) is that participants in Experiment 2 were less balanced bilinguals than participants in Experiments 1a/1b, but that

---

[51] The apparently greater language-switch cost for Experiment 1a's HF subset compared to Experiment 1a's full data set is presumably related to the fact that the data set was rather small in the former and was thus presumably more affected by variability.

the language tests employed were not sufficiently sensitive to ascertain this.

3.5 Conclusion

Experiment 2 replicated the trial type x task interaction shown in Chapter 2. The data pattern is compatible both with the external and internal, but not the input-switch or associative connections accounts. Experiment 2 furthermore showed that task order can affect the language-switch cost, but, it would seem, only in the language-specific task and only in the 'pure-HF' condition. This data pattern was predicted only by the external account. This finding is in accord with the suggestion (discussed in the Introduction to this chapter), that bilinguals presented with 'pure-HF' lists place much weight on wordness signals, contrary to individuals presented with 'mixed' lists. It seems that the large switch-cost associated with having to switch between task schemata can be reduced when one is able to place much weight on paralexical wordness signals that can bias towards the correct response. In the overall analysis (based on participants from all three groups, and thus the most powerful), Experiment 2 again showed smaller switch-costs in the language-general task than in the language-specific task, for both words and nonwords. Word and nonword data appear to differ with one respect though. In the language-general task there appears to still be a language-switch cost for words (analysis over all participants). For nonwords, in contrast, there was no statistically reliable evidence for a language-switch cost. Such a disparate pattern is in accord with the external account's task schema notion. The internal account finds such a disparity somewhat problematic.

In sum, the account best able to predict the data patterns (word and nonword) found here is the external account, while the internal account only predicts aspects of the obtained pattern. The input-switch account was not in accord with a large proportion of the data.

3.6 Outlook to Experiment 3

So far the language-switch cost can be accommodated by the external account. When decisions are language-specific, there are competing task schemata with strong mutual inhibition. Switching from one to the other gives rise to a substantial cost. When decisions are language-general, there is no need for strong (or even any) inhibition between schemata. Following Monsell, Doyle & Haggard (1989; but see also Grainger & Jacobs, 1996), it is supposed that signals of all sorts feed into word processing in an LDT. When language signals arise in a language-general task (in which language is not relevant), how do the bilingual participants deal with them? It was suggested that detecting a change of input language is interpreted by bilinguals participating in an LDT as meaning that the response for the current item needs to be different from the response for the previous item. What evidence is there for this? Given a response-repetition which coincides with a language-switch (as was the case in the HF subset of Experiment 1b and the critical HF items of Experiment 2, language-general task), that means that the data feeding into the analysis of trial type were only cases where a 'language-change' signal activated the wrong response units (and indeed, a significant RT delay was found). If one wants to be able to claim with some certainty 1) that 'language-change' signals play this (paralexical) role of activating the units underlying one response rather than the other, and 2) that bilinguals interpret a change of language signal as indicating that a change of response is required, then one needs also to demonstrate that a change of language signal can lead to a *faster* response in the case that a language-switch and a response-switch coincide. Such an outcome would be consistent with the external, but not the internal account. The findings of Thomas & Allport (2000, Experiment 1), using language-switching, as well as the findings of Rogers & Monsell (1995), using task-switching, suggest that such a pattern is possible. Experiment 3, in which language-switching and response-switching are fully crossed, follows up this conjecture.

Furthermore, to assess the role of 'language-change' signals - and thus distinguish between the external and internal accounts - one needs to attempt to divide out the contribution to language-switch costs (in a language-general task) that could stem from paralexical decision processes as opposed to stemming from task schema competition. Following de Groot and others, it is assumed here that semantic representations are common to both languages[52], in contrast to lexical representations (see also Dijkstra, van Jaarsfeld & ten Brinke, 1998, Figure 1, for the same assumption for the BIA model). In the language-general LDT it was assumed that individuals would establish a task schema for each language out of the habit of treating the languages differently at the lexical level. If meaning representations are indeed language-neutral, then there is less reason that, for normal purposes, bilinguals should treat processes relating to meaning differently for each language.

---

[52] de Groot, 1992a, 1992b, 1993. See Kroll & de Groot (1997) and de Groot (1998) for recent reviews of the relevant work. See Illes, Francis, Desmond, Gabrieli, Glover, Poldrack, Lee & Wagner, 1999, for fMRI confirmation of the behavioural findings.

Now, given a task that targets mainly semantic representations[53], there is a strong likelihood that bilinguals will establish only a single task schema for a semantic task, even if both languages should be presented overtly, but are not differentiated between by the task instructions (i.e., a language-general semantic task using the alternating runs paradigm). In sum, to increase the possibility of only a single task schema being set up (so as to be able to assess the effect of 'language-change' signals), Experiment 3 is a semantic categorisation task (specifically: an animacy decision task), which uses the alternating runs paradigm, and requires language-general decisions.

---

[53] Note: though the task only requires assessing semantic properties, part of the process of accessing them will be to activate lexical representations, and so properties relating to representations at the lexical level could impinge on the task which in itself targets the semantic level.

IV

Experiment 3: Using bilingual animacy decisions[54] to contrast an internal and an external account of language-switch costs

4.1 Introduction

Experiment 3 tests the external account's conjectures about the role of language-change signals in a language-general task. The language-switch cost in the language-general LDT was taken (on the external account) to reflect two task schemata, either weakly inhibited or jointly active. Additionally, language information was suggested to feed into the decision process paralexically, with language-change signals biasing towards the response previously not used. On a language-switch / response-repetition trial, this means that the wrong response units are activated. At the end of the preceding chapter it was said that, in order to be able to claim with some certainty that this is how such signals are interpreted, it is necessary to show that, given the right circumstances (i.e., where the language-change signals can bias towards the correct response units) a language-switch trial could show *faster* RT than a language-non-switch trial. This sort of finding would speak against an internal account, which predicts a language-switch cost whenever there is a language-switch. In short, to be able to distinguish between the external and internal accounts, it is necessary to attempt to create circumstances that would allow just a single task schema being set up, because when two task schemata are employed, the external account would also predict a language-switch cost (due to having to switch between the schemata). As discussed in the previous chapters, it may well be that a task which targets lexical representations is likely to induce two task schemata being set up, regardless of task instructions, while a semantic task may allow just a single schema to be established.

Experiment 3 fulfils various objectives:

(1) Experiment 3 tests the external and internal accounts further. The internal account predicts that there is a language-switch cost whenever there is a language-switch, because of the activation decrement of the 'new' language's lexical representations. In contrast, the external account predicts that language-switching incurs a cost either if there are separate

---

[54] "does the current word refer to an animate or an inanimate entity?" The meaning of the words 'animate' and 'inanimate' in the context of this thesis is 'living' and 'non-living', respectively.

task schemata (that need to be switched between), or because language-change signals (that feed into the decision process paralexically) bias towards the incorrect response units. That is, the external account suggests that it is not automatically the case that all language-switch RTs are longer than language-non-switch RTs. The means of differentiating between the two accounts lies in manipulating the role that language-change signals can play. This is described next.

(2) If language-change signals feeding into the decision process paralexically are interpreted as "a different response is needed now", and hence bias towards the units underlying the response previously not used, then this should only incur a cost on trials where the wrong response units have been biased towards. That is, on language-switch / response-repetition trials. If the bias has been towards the correct response units (as should be the case on, for example, language-switch / response-switch trials), there should be no 'language-switch cost'. That is, a language-switch trial need not necessarily be associated with a switch-cost. Rather, a language-switch cost should be modified by the type of response. The external account predicts an interaction between language-switching and response-switching (i.e., a language-switch cost for response-repetition trials, no language-switch cost for response-switch trials). To this end, it is necessary to fully cross language-switching and response-switching (which was not done before), as well as increase the chance of just a single task schema being set up (by using a semantic task and language-general task instructions).

(3) Experiment 3 is a semantic task (specifically: an animacy decision task, ADT). This fulfils two purposes. Firstly, it allows one to investigate whether, when the task focuses on mental representations that may be language-neutral[55], it is possible for a single task schema to be set up. Secondly, whether in that case language signals are still able to

---

[55] de Groot (e.g., 1992a, 1992b, 1993) proposes a system where shared meanings share representations at the semantic level. The distributed representations version of her proposal suggests that several nodes together make up the semantic representation of a word (each contributing an element of the meaning, e.g., four legs, furry, barks, domesticated, etc.). This approach seems a sensible and parsimonious approach, given that it avoids representing things more than once, but at the same time allows differences to be accounted for (e.g., a Frenchman will have somewhat different meaning referents for the word 'bread' than a German). Thus, the greater the meaning-overlap, the more nodes will be shared. The relevant nodes are then connected to the lexical representations, one for each language, reflecting the different word forms in each language (e.g., BREAD and BROT ['bread' in German]). Thus, at the semantic level there could be a common representational system, which does not have separate representations between languages and so does not 'code language' inherently, as is the case for lexical representations. However, by virtue of being connected to lexical representations that are language-coded, meaning representations and language information may interact (see de Groot, Delmaar & Lupker, 2000, p. 425). See Kroll & de Groot (1997) and de Groot (1998) for recent reviews of the relevant work. See Illes et al., 1999, for confirmation of such cognitive research, by use of fMRI.

influence processing. It is assumed that they can, since, even though processing focuses mainly on language-neutral semantic representations, activation of lexical representations is a part of the mental process. Activating the lexical representations will give rise to language signals (e.g., signalling that the lexical form APPLE of the animate entity {apple} is an English word). This would be a 'late language signal' An 'early language signal' could arise when the word's orthography is specific to one language only (e.g., SNOW for English, SCHNEE ['snow'] for German). How these language signals are then dealt with, is the question (as addressed in section (2) above).

In sum, in an animacy decision task, which presents words in English and German (alternating runs paradigm) and requires animacy decisions, an internal account predicts a language-switch cost regardless of response type, an external account predicts response type to affect the language-switching.

There are various types of semantic tasks. The choice for the animacy decision task was motivated as follows.

4.1.1 Animacy decisions

The focus of this thesis being on language-switching, the alternating-runs paradigm was to be used here as well, suggesting the use of a binary-response task (here: animate/inanimate [living/non-living]; in the previous chapter: word/nonword). One advantage of this type of task is that the same decision is utilised throughout. The other type of semantic task frequently found in the literature, category/instance decisions, requires making decisions where the target that a word is matched against changes constantly. For example: "is APPLE an instance of the category FRUIT? ... is TABLE an instance of the category VEGETABLE? ... is ROSE an instance of the category FURNITURE? ...". In a *bilingual* category/instance task, the design for crossing language-switching and category-switching would be very complex, requiring many items, which may be difficult to obtain and which would make a task inordinately long. In contrast, animacy decisions require the same decisions throughout. There is an additional advantage to animacy decisions. They allow decisions on each item (rather than on every other item, which is necessarily the case when category labels and category instances need to be presented), and for a single word at a time (rather than word pairs). For these reasons, animacy decisions seem better suited to an on-line investigation of language-switching effects. Furthermore, while category/instance decisions may be subject to individual variation (e.g., is TOMATO a FRUIT or a VEGETABLE?), 'animacy' seems to be a very fundamental category. Converging evidence for this position has come from cognitive behavioural studies with young healthy infants (e.g., Spelke, Breinlinger, Macomber & Jacobson, 1992), neuroimaging (PET scan) studies (e.g., Perani, Schnur, Tettamanti, Gorno-Tempini, Cappa & Fazio, 1999), as well as clinical cases (e.g., Warrington & Shallice, 1984; Sartory & Job, 1988; Farah, McMullen & Meyer, 1991). Perani et al.'s study indicated that the recognition of animate and inanimate entities is subserved by different brain networks, while clinical studies have shown that it is possible for mental representations of animate and inanimate entities to be impaired selectively, such that, for example, individuals are able to name inanimate items, but not animate ones - or vice versa (e.g., Warrington & Shallice, 1984).

Such findings suggest that animacy is a very basic type of concept, likely to be language-independent. This serves the purpose of the present experiment, since 1) individuals carry out a task in which their judgements do not require evaluating representations at a level differentiating between languages, 2) the distinctions to be made are based on broad categories, and 3) decisions are binary, i.e. are unchanged throughout the task.

4.1.2 Bilingual animacy decisions and their use for differentiating between the internal and external accounts of control

In a bilingual version of the animacy decision task the words presented are in one of two languages (e.g., presented singly, in succession: APPLE, TISCH[56], BLUME[57], CARROT, ...). In such a task, language is clearly irrelevant to the decision, but the two languages (and the switches from one to the other) are plainly present. If a language-switch cost was found in a language-general LDT, where language is also irrelevant, does this entail that there would also be a language-switch cost in a bilingual animacy decision task (ADT), for which language is also immaterial? Not necessarily. What needs to be considered is 1) whether or not lexical representations of the non-target language are inhibited during processing in the target language (as the internal account suggests), 2) what role language signals and 'language-change' signals play, and 3) what representations the task targets (lexical vs.

---

[56] 'table' in German

[57] 'flower' in German

semantic). These are each discussed further in the following.

4.1.3 Predictions

The two models tested by this experiment are the internal account and the external account. The predictions made by these two accounts with regard to a language-switch cost in a language-general ADT are discussed next.

Internal account: Assume that a language A word (e.g., 'DOG') is presented to a bilingual in a language-general ADT. The word is analysed and identified as a lexical representation with a link to language-node A. Part of this process is that (both via top-down inhibition from language-node A and via lateral inhibitory links at the lexical level), language B lexical representations are suppressed. The meaning ({dog}) is accessed, an animacy decision is taken, and a response is given. On the following trial a word in language B is presented (e.g., 'ULME', 'elm' in German). The activation of language-node A does not return to zero automatically and immediately (Dijkstra & van Heuven, 1998), and so could continue to inhibit language B lexical representations for a while. Thus, when a language B item is presented on a language-switch trial, language B lexical representations are disadvantaged vis-à-vis language A lexical representations. Overcoming this activation deficit takes time, and so RT will be slower on language-switch compared to language-non-switch trials. If task instructions do not specify making the decisions in a specific language (i.e., it is a language-general, not a language-specific, task), then the cost for switching may be small, because top-down inhibition is relatively weak (as it was in the language-general LDT), but there will be a cost, since a language-switch, with the accompanying need to overcome an activation decrement, does take place.

External account: Taking the same scenario as above, assume that the word 'DOG' is presented. It is identified and meaning is accessed. In a task focusing on apparently language-neutral meaning representations (rather than language-coded lexical representations), and given language-general task instructions, it is conceivable that only a single task schema is set up. Such a single task schema (e.g., "activate output 'A' if animate, activate output 'I' if inanimate") links the signal from the lexico-semantic system to the response mechanism (in the case of the word 'DOG', to response 'A'). The units for response 'A' are in competition with those of response 'I'. In order for 'A' to capture the response, it needs to exceed the activation of response 'I' (see Levelt, Roelofs & Meyer, 1999, for a relevant competition-sensitive mechanism). On the following trial the word 'ULME' (elm) is presented. The features are analysed, the word-form representation is activated, meaning is accessed. This word too is animate (as 'DOG' before it was), and so the same response is employed again ('A'). So far, on this account the RTs for 'DOG' and 'ULME' should not differ (unlike in the internal account), because no reconfiguration was required (the same response could be re-employed and one task schema is used throughout). In other words, even though there is a shift in language there should not be a cost for doing so. If instead of 'ULME' the word 'TISCH' ('table' in German) were presented after 'DOG', the response 'I' would need to become more active than the previously superior 'A'. This would take time and so RT for 'TISCH' would be delayed relative to that for 'DOG'. Note that, as discussed so far, the external account would also predict a delay for 'TABLE', if following 'DOG' (a 'response-switch' cost). I.e., a delay may be expected when response-reconfiguration is required (cf. Pashler & Baylis, 1991; Rabbit, 1968). There should not be a cost for language-switching per se. In the Introduction to this chapter it was noted briefly that language still plays a role, via the 'paralexical decision processes'. Their influence will affect the scenario presented so far.

4.1.4 The role of language signals and 'language-change' signals

Recall from previous chapters the discussion about the multitude of signals available for and utilised in making decisions ('paralexical decision processes', Monsell, Doyle & Haggard, 1989; 'multiple read-out' Grainger & Jacobs, 1996). Presenting individuals with words from both languages (alternating runs paradigm) and asking them to make animacy decisions, does not require recourse to language information. However, individuals will presumably register the words' language (cf. Kroll & Dijkstra, 2001) and hence language-switches. One of the several signals feeding into the paralexical decision process are likely to be language signals. A change in language could be noted as "a change has occurred", suggesting that the current item differs from the preceding one (as discussed in Chapter 1). This could result in biasing towards the response which was *not* selected on the previous trial. Biasing towards the formerly non-selected response units is advantageous for the next reply if a change is indeed required. If, after all, no change is required, then the just-used response units will need to be activated sufficiently above the ones incorrectly biased towards by the paralexical decision process. 'Revoking the bias' will take time, manifested in delayed RT. What the proposal for the ADT is, when taking account of paralexical decision processes, is now discussed for the external and internal accounts.

The *external account* predicts that there may be an overall language-switch cost, but that there will be an interaction between language-switching and response-switching (i.e., animacy-switching). A similar pattern to that predicted here was found by Thomas & Allport (2000): In a bilingual LDT participants reacted more slowly on language-switch than on language-non-switch trials only for repeated responses; when there was a response change there were no language-switch costs. (See also Rogers & Monsell, 1995, for a task-switching experiment with the corresponding pattern.) The pattern of a normal language-switch cost should be found when the response remains unchanged (response-repetition trials); i.e., delayed RTs on language-switch trials compared to language-non-switch trials, because language-change signals bias towards an incorrect response. However, when the response needs to change (response-switch trials), language-change signals would bias towards the correct response, in which case language-switch RTs could be *faster* than language-non-switch RTs.

There are two cases where language-switching and response-switching cross (language-non-switch / response-switch and language-switch / response-repetition). As discussed above, a delay may be expected for both these cases. How great the delay on each of these two types of trials is, can be speculated on as follows. Monsell et al. (1989) suggest that paralexical decision processes take place while the main processing for the stimulus is in progress. As soon as relevant signals are available, they begin to feed into the calculation of what the next response is most likely to be. Presumably, the sooner such signals are available, the sooner they can begin biasing towards the next response (or the units underlying it), achieving a stronger bias than would be the case if such signals were available later. Should language information be available before semantic information, then, on a language-switch/response-repetition trial, the signal pertaining to the language-switch would begin biasing towards the response at an early stage, allowing a strong bias to be obtained before the attribute (animate or inanimate) has been determined[58]. For a language-non-switch/response-switch trial the argument would be that language indicates that "no change is necessary",

---

[58] An orthographically based language signal (e.g., SCHNEE or SNOW, orthographically legal only in their respective languages) would be available early, and is likely to be available before animacy information has been retrieved, since this signal should become available at a very peripheral level of language processing. Language information based on a language tag or node would be available comparatively later (being reliant on lexical access), but could nonetheless still be available before semantic information.

upon which the previously used response is biased towards, when in fact the response previously not used will be required[59].

If the signals "something has changed" and "nothing has changed" are equally strong, then the delay on language-switch / response-repetition trials and language-non-switch / response-switch trials should be comparable. If, however, a 'positive' signal ("something has changed") is stronger than a 'negative' signal ("nothing has changed"), then the delay on language-switch / response-repetition trials should be greater than that on language-non-switch / response-switch trials. This is, however, only a subsidiary point. The focus of Experiment 3 is on the language-switching cost and the influence of response type on language-switching costs, in order to differentiate between the internal and external accounts, and to assess the role of 'language-change' signals.

Note, in this task, language signals provide a correct cue in only 50% of the cases (on language-switch/response-switch and on language-non-switch/response-repetition trials). In other words, signals regarding language-change / no-change are not a reliably predictive cue for responses, and so, one might argue, individuals should not be heeding them. However, if language information is a factor in normal, everyday, bilingual language functioning, then it may be that it cannot easily be 'turned off' just for the benefit of a half-hour experiment. At the same time, it is possible that individuals eventually reduce the weight attached to language-change signals paralexically, and so avoid being delayed by the misleading bias. If that is possible, then one might find that, after some period of task experience, there is no longer a delay on language-switch/response-repetition trials (the ones that experience the incorrect bias initially).

An *internal account* with top-down inhibition (language-node to lexical level) predicts a language-switch cost each time language changes, because each time this happens an activation deficit needs to be overcome. Adding paralexical processes into the equation could possibly permit a language-switching x response-switching interaction. Language signals available early on (e.g., from language-specific orthography) could bias either towards the

---

[59] As an alternative it is conceivable that there is no additional activation when there is no change. But even in this case there would be an advantage of the response units used on the previous trial (compared to the ones not used), as there could well still be residual activation from the previous trial. ('Residual activation from the previous trial' is a notion that returns in the Chapter 5, in the 'carry-over effects'.)

previously used response (in the case of 'language unchanged' signals), or towards the response previously not used (in the case of 'language has changed' signals). This would be occurring while an activation decrement is being overcome. Once that is achieved and the relevant semantic attribute has been ascertained, activation can be fed to the corresponding response unit. If that response unit has been preactivated, it could be faster to take it above threshold. In sum, the 'natural' language-switch cost is exaggerated on response-repetition trials (due to incorrect bias), but it is decreased on response-switch trials (due to correct bias), providing an interaction between language-switching and response-switching. However, unlike the external account, learning to reduce the effect of language-change signals does not eliminate a language-switch cost. It would only mean decreasing it down to the 'natural' language-switch cost (i.e., without the exaggeration caused by the incorrect bias).

If the BIA model allowed top-down inhibition (language-node to lexical level) to not only be weakened, given task demands (i.e., language-general vs. language-specific LDT), but to even be reduced to zero, then there would be no inhibition of other-language representations and hence no activation decrement - and no language-switch cost. The BIA literature has not, however, proposed this possibility.

4.1.5 Representations targeted by the animacy decision task

An animacy decision task requires output from the semantic level, where animacy could be coded as "+animate" / "-animate" [living/non-living] 'attached to' meaning representations, or it could be computed from a set of attributes (e.g., 'it grows, requires nourishment, can be diseased, will die' etc.). Bilingual semantic studies to be found in the literature aimed to establish whether bilinguals have separate or shared representations. They addressed this question mainly with category/instance tasks, in which the category names and instances were either of the same language (e.g., FRUIT/APPLE) or different languages (e.g., FRUIT/POMME ['apple' in French]). Irrespective of whether presentation of category names and instances was simultaneous or temporally separated, there was no indication that decisions were any longer for 'mixed language' than for 'same language' pairs (e.g., Caramazza & Brones, 1980; Shanon, 1982; Potter, So, von Eckhardt & Feldman, 1984; Dufour & Kroll, 1995[60]). The same result was obtained in a synonym judgement task with

---

[60] The exception to this finding (RT for mixed-language pair = RT for same-language pair) were novice bilinguals in Dufour & Kroll's (1995) category/instance task, whose decisions on mixed-language trials were delayed relative to those on same-language trials.

mixed- and same-language pairs (McCormack, 1977). None of these studies directly addressed the question of language-switching, and they do not provide information on the direction or pattern of language-switches. Though these tasks are not ideally suited to addressing language-switching in bilinguals, they nonetheless suggest that there is no cost for language-switching in a semantic task for bilinguals who are presented with stimuli from each of their languages. If there were a language-switch cost whenever there is a language-switch (after all, both the category label and the category instance need to be processed), then mixed-language-pair decisions should have taken longer than same-language-pair decisions, since the former involve a language-switch but the latter do not.

To address the effect of language-switching directly, a task permitting decisions on every trial is preferable. The only existing study that could be identified is an unpublished undergraduate project (Mason, 1994). This animacy decision task, using the alternating runs paradigm, showed no significant effect of language-switching (mean RT on language-switch trials: 549ms, on language-non-switch trials: 550ms). This study did not, however, manipulate typicality (to ascertain sensitivity). More importantly, the effect of language-switching seems to have been investigated without taking account of the decisions taken (response-switch/response-repetition) and whether they coincided with language-switches or not. As discussed previously, paralexical decision processes influence decisions. Thus, conceivably, even though language is irrelevant for the decision, it may feed into the decision nonetheless and hence influence response time.

In sum, it is predicted here (following the external account) that there will be no language-switch cost per se. If 'language-change' signals are indeed interpreted by bilingual participants to mean that their response should change, then there should be an interaction between language-switching and response-switching. Due to language-change signals feeding into the decision process paralexically and biasing towards certain response units, language-switch trials are predicted to have slower RTs than language-non-switch trials when they coincide with response-repetition. When there is a change of response, there should be no such language-switch cost. The normally found response-repetition effect may reverse (overall), since language-change signals bias towards the wrong response on language-switch / response-repetition trials. The *internal* account, in contrast, predicts a significant (if small) language-switch cost, presumably comparable to that in the language-general LDT. But in any event, it is unable to predict that the language-switch cost should be present only for response-repetition trials. Furthermore, this account is unable to predict faster RTs for language-switch than language-non-switch when response-switching, as the external account does. An alternative internal account is possible. On the supposition that individuals in the present experiment enter a bilingual mode in which representations in both languages are equally active (Grosjean, 1997; 2001), there may be no cost in language-switching - repeated responses will simply be faster than non-repeated responses.

4.2 Method

4.2.1 Participants

Twenty-four (right-handed) German/English bilingual students from University College London (12 males, 12 females; mean age 26.2 (4.7) years, range 20-36), took part in this experiment. All had normal or corrected vision. Participation was voluntary and remunerated at the standard departmental rate.

Language proficiency:

All participants grew up with German and learned English as a second language at school, starting at a mean age of 10.1 (2.3) years. They had become fluent in English at a mean age of 17.8 (6.9) years (range: 13-27) and reported to have used English predominantly for the past 5.8 (5.1) years (range: 1-18), with the majority around 4 years.

Following the experiment, participants carried out the LLEX tests Meara, 1994; for details see Chapter 2) and completed the language background questionnaire (see Appendix 1). In the LLEX tests (maximum scores: 100 per test) results ranged 79-98 (mean = 90.8, *SD* = 5.8) for English, and 65-98 (mean = 91.0, *SD* = 7.8) for German, suggesting that these bilinguals were reasonably well balanced. However, the result of the self-rating (self-assessing their relative proficiency in German compared to English, out of a combined total of 100 [see the language-background questionnaire in Appendix 1]), indicated somewhat higher estimates for German than for English (German: mean = 58.3, *SD* = 7.4). Participants reported their current use of the two languages (again out of a combined total of 100) to be predominantly English (mean = 76.5, *SD* = 13.6). Fifteen participants chose the German, nine the English version of the questionnaire.

4.2.2 Design

The repeated measures design comprised four factors each with two levels: language (English / German), language-switching (language-switch / language-non-switch), animacy (animate / inanimate) and response type (response-switch [animacy-switch] / response-repetition [animacy-non-switch]). To permit an investigation of any changes over time (e.g., concerning how much weight participants attached to language-change signals) there was an additional factor: part (part 1 / part 2), referring to the first and second half of the experiment.

The animate and inanimate categories each comprised eight subcategories (with four typical and four less typical exemplars in each), yielding a basic item set of 128 items. Language switches occurred on alternating trials as per the alternating runs paradigm (Rogers & Monsell, 1995). A filler trial at the start of each block of 64 trials designated the first measured item as a switch trial or a non-switch trial. The first two blocks ('part 1') of trials comprised the full item list of 128 items. The next two blocks ('part 2') were the translations of the words seen in the first two blocks. Words that appeared in blocks 3 and 4 for one half of the participants were presented in blocks 1 and 2 for the other half; translations were never in immediately adjacent blocks.

Words that appeared on language-switch trials for one half of the participants, were presented on language-non-switch trials for the other half. Likewise, words that appeared on animacy-switch trials (response-switch) for one half of the participants were presented on animacy-non-switch trials (response-repetition) for the other half. Adjacent items were from different subcategories and were not obvious associates. Over the entire experiment each participant was shown the full set of materials in both English and in German, and over all participants each item, by its English name and by its German name, was presented on a language-switch and language-non-switch trial in the context of a repeated or a non-repeated response.

4.2.3 Materials

The sets of 64 animate and 64 inanimate items were selected using the Battig & Montague (1969) norms – see Appendix 10 for a listing. To have a means of assessing this experiment's sensitivity to semantic factors, the degree of typicality was varied. (Note that this experiment was not aimed at investigating typicality. Typicality was varied in order to have a means of assessing sensitivity to semantic properties.) The four typical instances in each subcategory generally had a frequency of mention of 100 or more, whereas the poor, or less typical, instances generally had a frequency of mention of less than 50. In all cases, there was always a gap of at least fifty points between the more and less typical exemplars within each subcategory. For English, the exemplars of each type for each subcategory were matched in frequency (Kučera & Francis, 1967) as well as in letter length and in syllable length - see Table 4.1. On average, German words were longer and of lower CELEX frequency, though they did not differ in terms of the rated animacy of their referents (see below).

Table 4.1
Mean values (SD in brackets) of semantic properties (category typicality, Battig & Montague, 1969; subjective animacy ratings[1]), word frequencies (Kučera & Francis, 1967, for English; Ruoff (1990), for German; CELEX[2] for English and German respectively) and word length (letters and syllables) for all word stimuli, as a function of language, animacy and typicality. B&M = Battig & Montague; K&F = Kučera & Francis.

Language	Animacy	Typicality	Semantic Properties		Word Frequency		Word Length	
			B & M Typicality Rating	Subjective Animacy Rating[1]	K & F	CELEX (E)	Letters	Syllables
English	**Animate**	**good**	201 (94)	7.2 (0.7)	5 (5)	6 (5)	5 (1)	2 (1)
		poor	25 (17)	6.9 (0.8)	6 (13)	4 (4)	5 (2)	2 (1)
	Inanimate	**good**	238 (105)	1.7 (0.4)	27 (33)	30 (47)	5 (1)	1 (1)
		poor	18 (15)	1.9 (1.0)	26 (33)	19 (21)	6 (1)	2 (1)
					Ruoff	**CELEX (G)**		
German	**Animate**	**good**	-	6.9 (1.0)	12 (27)	2 (2)	6 (2)	2 (1)
		poor	-	6.7 (1.0)	12 (22)	2 (3)	6 (2)	2 (1)
	Inanimate	**good**	-	1.9 (0.6)	15 (26)	12 (19)	6 (1)	2 (1)
		poor	-	2.0 (0.9)	15 (19)	15 (41)	6 (1)	2 (1)

[1] Ratings were given on an 8-point scale (8 = animate, 1 = inanimate); for English items they were given by English unilinguals, for German items by German/English bilinguals - details are given in the following section.

[2] Note: items were matched (within each subcategory) using Kučera & Francis (1967) norms; Ruoff (1990) and CELEX values are given for comparative purposes - CELEX data were available only after the experiment was completed, thus frequency matching was based on Kučera & Francis (1967) norms.

Animacy rating study:

To ensure that all items fell clearly into either the animate or inanimate category (and that this was comparable for German and English word meanings), an animacy ('livingness') rating study was undertaken.

Twenty-two English unilinguals (undergraduates in the Department of Psychology, University College London) received randomised lists of the word stimuli and rated the referents of each word on an 8-point animacy scale, with the end points, suitably rotated over participants, marked as 'animate' and as 'inanimate'. A further twenty-two German/English bilinguals living in London, and from a variety of academic backgrounds, performed the same task in German - none participated in the following on-line study. Individuals were told to rate each word with respect to how good an example it was of 'living' or 'non-living' entities, using the 8-point scale. For English raters, the average rating for animate items (with higher values indicating higher rated animacy) was 7.1 (0.8). For inanimate items, the average rating was 1.8 (0.8). German/English bilingual raters showed the same disjoint pattern: mean (animate) = 6.9 (1.0), mean (inanimate) = 1.9 (0.8). Ratings showed no reliable effect of typicality for either group (one-way ANOVAs yielded Fs < 1 or *ps* > 0.1), indicating that typicality did not interfere with the crucial animacy categories.

4.2.4 Apparatus & Procedure

Participants, tested individually in a small cubicle using an IBM compatible PC, were given verbal instructions in German followed by written instructions in English. This was done to ensure that representations of both languages were active before starting the experiment (if one language were in a comparatively less active state from lack of use, individuals might be slower in identifying word representations of that language, and hence be delayed with their animacy decisions). Participants were informed that they would be shown words, some English and some German, all of which were the names of living or non-living entities. Their task was to decide as quickly and as accurately as possible whether the entity named was living (animate) or non-living (inanimate). (Note, verbal instructions used the words 'lebend' [living] and 'nicht-lebend' [non-living], so that participants would have been clear about the intended distinction). A two-button response box registered responses. Participants pressed the '+' button ('animate') with the middle finger of their dominant hand, and the '-' button ('inanimate') with the index finger of the same hand. The index finger of the dominant hand is likely to be faster at button-pressing, not least due to practice of this action with a PC mouse's left button. It would have been generally preferable to have balanced finger assignment over participants, but assigning the usually slower category (inanimate) to the 'better' finger (index finger) goes against the pattern reported in the literature (faster responses for animate). In any event, this experiment did not aim to ascertain an animacy effect, so that this short-coming is not considered too grave.

Response times were measured from the onset of the word's presentation to a button press. Errors elicited a short beep. One second after the button-press the next word appeared on the screen and remained on the screen until terminated by a response.

Words appeared singly in white uppercase letters centred on a (black) computer screen. The uppercase format allowed the same visual presentation for English and German words, given the need to capitalise the initial letter of German nouns. To create a 'language-general' task (allowing individuals to establish a single task schema), participants were told that language was irrelevant to their task. Following (for instance) Kroll & Dijkstra (2001), it was assumed that language signals would arise during processing, irrespective of being told that language was irrelevant, and that participants would be influenced by them. Participants were also told that they might notice translations (i.e., that a word would be shown in one language at one point and in the other language at some other point), but that this should not concern them. This was done to stop participants from being surprised at seeing translations, and possibly thinking about this, as their RTs might then be delayed.

After a practice block of 16 trials, none of which appeared in the experimental trials, individuals paced themselves through the experimental trials, counting slowly to twenty between blocks, before initiating the next block of trials. The experiment took on average 20 minutes to complete.

After finishing the experiment, participants carried out the two LLEX tests (Meara, 1994) and filled in the language background questionnaire. The LLEX tests required an additional 30 minutes. Language background questionnaires were completed at home. All participants returned their questionnaires to the experimenter at the latest a week later.

4.3 Results & Discussion

The results on RTs and error rates reported here stem from different analyses: (1) on all data, including the additional factor 'part' and (2) from analyses for parts 1 and 2 individually.

In the 'overall' analyses there were five factors: Part (part 1 / part 2), Language (English / German), Language-Switching (language-switch / language-non-switch), Animacy (animate / inanimate) and Response Type (response-repetition [animacy-non-switch] / response-switch [animacy-switch]). The individual analyses for parts 1 and 2 had the same factors as the overall analyses, excluding the factor Part. In the item analyses Animacy was a between-subjects factor.

4.3.1 Data treatment

The RT data excluded latencies +/- 2 standard deviations (*SD*) from each participant's cell mean (a total of 4.0% trials of the whole data set)[61]. Errors totalled 7.7%; error analyses were based on the arcsine transformed error proportions. Analyses were carried out both on mean and median correct RTs (for median analyses only errors, not outliers, were excluded). The two analyses followed the same pattern. The data reported here are the analyses on the mean correct RTs, from which outliers are excluded[62]. In this way the data reported here are most compatible with the data presentation in Experiments 1a/1b and 2. Inspection of the RT and error data revealed no speed/accuracy trade-off. Mean values presented in the text and tables are mean of means (from the correct subject data). Switch-costs presented here were calculated by subtracting the data (RT / error) of the non-switch trial from that of the switch trial.

4.3.2 The main RT and error analyses

In the following, results from the overall and separate analyses are not strictly separated, to avoid dealing with each topic more than once.

First, the main effects arising in the overall analysis are presented. Where relevant, the separate analyses for parts 1 and 2 are provided as well (for instance to show that an effect significant in part 1 ceased being so in part 2). This is followed by any interactions arising in the overall analyses, again with data from parts 1 and part 2 reported separately where relevant. Against this background, the critical language-switching and response-type data, for

---

[61] For trials in English, a coding error, discovered retrospectively, decreased the size of two of the cells (animacy-switch, for both animate and inanimate items) and enlarged two others (animacy-non-switch, for both animate and inanimate items). Each participant should have received each trial type four times per block. In the reduced cases they received it three times and in the increased cases five times per block. The smallest cell size still yielded 12 instances over the entire experiment. A further analysis, using the same factors as the reported analysis but in which data points were (randomly) eliminated from all complete cells (rendering all cells the same size), showed very similar patterns. As a further check, analyses were carried out on the full German set alone (which was free of coding errors) and on the full (unbalanced) English set alone. The German and English analyses showed very similar patters. (See Appendix 7 for these four analyses.) On these grounds it was decided that the full analyses could be pursued.

[62] Analyses were also carried out on mean and median correct RTs from which RTs following each wrong response were excluded. These followed the same data patterns as the analyses excluding only errors themselves. They are therefore not reported here either.

parts 1 and 2 respectively, are then presented.

Overall, RTs were slower in part 1 than in part 2 [852ms (154) vs. 804ms (151); $F_1(1,23) = 25.44$, MSE = 17767.00, $p < 0.001$; $F_2(1,126) = 117.91$, MSE = 12203,85, $p < 0.001$; min F'(1,34) = 20.93, $p < 0.001$], but there was no difference between the two parts in terms of error rates [$F_1(1,23) = 1.96, p > 0.1$; $F_2 < 1$; min F' < 1].

As might be expected, given their self-ratings, participants responded more quickly to German words than to English words [800ms (142) vs. 856ms (162)], though this was significant only by subjects [$F_1(1,23) = 35.70$, MSE = 17071.94, $p < 0.001$; $F_2(1,126) = 1.29$, $p > 0.2$; min F'(1,134) = 1.25, $p > 0.2$]. Participants also made fewer errors on German words [4.7% vs. 10.8%; $F_1(1,23) = 30.79$, MSE = 0.11, $p < 0.001$; $F_2(1,126) = 26.20$, MSE = 0.02, $p < 0.001$; min F'(1,87) = 14.16, $p < 0.001$]. None of these effects were modified by the factor part [$p > 0.1$ or $F < 1$].

In line with the literature (e.g., de Groot, 1990), participants reached an animate decision more quickly than an inanimate decision [798ms (150) vs. 858ms (153)], though this was significant only by subjects [$F_1(1,23) = 35.64$, MSE = 19166.20, $p < 0.001$; $F_2 < 1$; min F' < 1]. In terms of errors, there was no difference between 'animate' and 'inanimate' responses [$p > 0.1$ both by subjects and by items, min F' < 1]. The RT difference was obtained despite words referring to inanimate items having a higher word frequency than items referring to animate items, and being allocated to the 'faster' finger. (But recall that it had not been the aim of this experiment to establish an animacy effect.) The main effect of animacy was modified by the factor part only in the subject RT analysis [$F_1(1,23) = 6.96$, MSE = 7827.27, $p < 0.025$; $F_2 < 1$; min F' < 1], in that the difference between animate and inanimate was greater in part 1 [814ms (154) vs. 890ms (144), respectively] than in part 2 [782ms (144) vs. 825ms (155), respectively]. For errors there was no such interaction [Fs < 1].

For the factor response type, RTs were, overall, 31ms *slower* on repeated trials compared to non-repeated trials [843ms (167) vs. 812ms (139); $F_1(1,23) = 19.13$, MSE = 9632.32, $p < 0.001$; $F_2(1,126) = 8.94$, MSE = 11021.67, $p < 0.01$; min F'(1,124) = 6.09, $p < 0.025$]. This pattern, which runs counter to the normally found advantage of response-repetition, can be understood if one allows other signals from within the system (specifically: language signals) to play a cueing role. This issue is pursued further below, when the predicted interaction between language-switching and response type is addressed. The main effect of response type was modified by the factor part by subjects and items, but only marginally by min F' [$F_1(1,23) = 5.25$, MSE = 4495.61, $p < 0.05$; $F_2(1,126) = 6.02$, MSE = 11667.70, $p < 0.025$; min F'(1,71) = 2.80, $p = 0.098$]. The effect of response type (repeated vs. non-repeated) was significant in part 1 [$F_1(1,23) = 25.45$, MSE = 6675.97, $p < 0.001$; $F_2(1,126) = 16.63$, MSE = 10077.02, $p < 0.001$; min F'(1,103) = 10.06, $p = 0.001$], whereas in part 2 it was significant only by subjects [$F_1(1,23) = 5.09$, MSE = 7451.97, $p < 0.05$; $F_2 < 1$; min F' < 1]. The difference between response-repetition and response-switch was 42ms in part 1 [mean RTs: 873ms (171) and 831ms (132) respectively], and only 20ms in part 2 [mean RTs: 814ms (157) and 794ms (144) respectively]. This suggests that whatever interfered with response generation in part 1, carried less weight in part 2; this is addressed further in the language-switching / response-type sections, below. For errors there was no significant difference between response-switch and response-repetition trials [$F_1(1,23) = 3.16$, MSE = 0.06, $p = 0.089$; F_2 (1,126) = 1.20, $p > 0.2$; min F' < 1], and this was not further modified by the factor part [$p > 0.1$ and F < 1].

The main effect of language-switching was significant in the overall analysis [$F_1(1,23) = 6.01$, MSE = 7441.73, $p < 0.025$; $F_2(1,126) = 14.34$, MSE = 15448.93, $p = 0.001$; min F'(1,45) = 4.24, $p < 0.05$]. Overall, RTs were slower on language-switch trials [836ms (165)] than on language-non-switch trials [820ms (142)]. This main effect was marginally modified by the factor part in the subject analysis [$F_1(1,23) = 3.44$, MSE = 6465.98, $p = 0.077$; $F_2 < 1$; min F' < 1]. In part 1 the difference between language-switch and language-non-switch RTs was a significant 26ms [mean RTs were 865ms (173) and 839ms (132) respectively; $F_1(1,23) = 7.49$, MSE = 8673.87, $p < 0.025$; $F_2(1,126) = 9.76$, MSE = 13465.21, $p < 0.01$; min F'(1,65) = 4.24, $p < 0.05$]. In part 2, however, the difference between language-switch and language-non-switch trials was only 5ms, and not significant [mean RTs were 806ms (152) and 801ms (150) respectively; Fs < 1]. In terms or errors, was no significant effect of language-switching [$p > 0.3$ and F < 1], and this was not modified by the factor part.

There were two two-way interactions for the RT data, that were, however, each significant only by subjects. For this reason not much significance is attached to them here, but they are reported for completeness' sake.

In terms of RT, language and animacy interacted, though this was significant only by subjects [$F_1(1,23) = 33.31$, MSE = 6314.97, $p < 0.001$; $F_2(1,126) = 1.08$, $p > 0.3$; min F'(1,134) = 1.05, $p > 0.3$]. The RT difference between animate and inanimate words was smaller for English words [mean RTs were, respectively, 843ms (163) and 869ms (159)] than for German words [mean RTs were, respectively, 753ms (119) and 846ms (147)]. This interaction was not further modified by part [Fs < 1]. Language and animacy also interacted in terms of errors [$F_1(1,23) = 17.36$, MSE = 0.07, $p < 0.001$; $F_2(1,126) = 9.56$, MSE = 0.02, $p < 0.01$; min F'(1,114) = 6.16, $p < 0.025$]. On English trials, there were clearly more errors for animate (14.2%) than for inanimate items (7.9%); significant by subjects and items, but only marginally by min F' [$F_1(1,23) = 5.10$, MSE = 0.03, $p < 0.05$; $F_2(1,126) = 13.88$, MSE = 0.09, $p = 0.001$; min F'(1,42) = 3.73, $p = 0.060$], whereas on German trials there were somewhat fewer errors for animate (3.8%) than for inanimate items (5.5%) [significant by items, marginal by subjects, n.s. by min F': $F_1(1,23) = 3.94$, MSE = 0.05, $p = 0.059$; $F_2(1,126) = 4.47$, MSE = 0.01, $p < 0.05$; min F'(1,71) = 2.09, $p > 0.1$]. Since word frequencies were lower for animate than for inanimate items, it is conceivable that the subjective frequency of English words referring to animate items was much lower for these bilingual participants[63]. Language x animacy was not further modified by the factor part [Fs < 1].

Animacy and response type interacted in the overall analysis, but only by subjects [$F_1(1,23) = 16.54$, MSE = 5770.00 $p < 0.001$; $F_2(1,126) = 1.48$, $p > 0.2$, min F'(1,143) = 1.36, $p > 0.2$]. This was not further modified by the factor part [$p > 0.1$ and F < 1]. Although participants were overall slower on response-repetition trials than on response-switch trials (see above), this pattern may be primarily attributable to inanimate items. For animate items, the means for response-switch and response-repetition trials were, respectively, 794ms (134) and 802ms (164) [significant by items but not by subjects or min F': $F_1(1,23) = 1.35$, $p > 0.2$; $F_2(1,63) = 8.14$, MSE = 11982.46, $p < 0.01$; min F'(1,31) = 1.16, $p > 0.2$]. In contrast, for inanimate items, the respective means were: 831ms (142) and 884ms (160) [significant by subjects but not by items or min F': $F_1(1,23) = 27.10$, MSE = 10053.79, $p < 0.001$; $F_2(1,63) = 1.72$, $p > 0.1$; min F'(1,139) = 1.62, $p > 0.2$]. There is no immediately plausible reason for

---

[63] To check that RTs were sensitive to the typicality of the items, an ANOVA was carried out with the factors language (English / German), language-switching (language-switch / language-non-switch), response type (response-repetition [animacy-non-switch] / response-switch [animacy-switch]) and typicality (typical / less typical), pooling over animate and inanimate items. Participants reacted more quickly to 'typical' than to 'atypical' items [805ms (140) vs. 842ms (138); $F_1(1,23) = 30.40$, MSE = 13492.14, $p < 0.001$; $F_2(1,126) = 6.68$, MSE = 13769.67, $p < 0.025$; min F'(1,148) = 5.48, $p < 0.025$] and made fewer errors on such items (6.1% vs. 10.2%), significant by subjects and items, but only marginally by min F' [$F_1(1,23) = 22.65$, MSE = 0.05, $p = 0.001$; $F_2(1,126) = 4.26$, MSE = 0.01, $p < 0.05$; min F'(1,149) = 3.59, $p = 0.060$]. Typicality interacted with none of the other factors [Fs < 1 or $ps > 0.05$]. Note, since typicality had not been envisaged as a separate analytical factor at the outset, cell sizes were very small, and with error and outlier deletion, some cells were reduced to zero. Hence for an analysis to be feasible, it was necessary to pool data, as was done here.

explaining this pattern; however, for both of the two-way interactions presented here, it should be noted that statistically there was never significance by both subjects and items, suggesting that the data are not very robust and may not be replicable. The corresponding error analyses were n.s. [*ps* > 0.1 or Fs < 1]. There were no further two-way interactions, in either the RT or in the error data (Fs <1 or *ps* > 0.1).

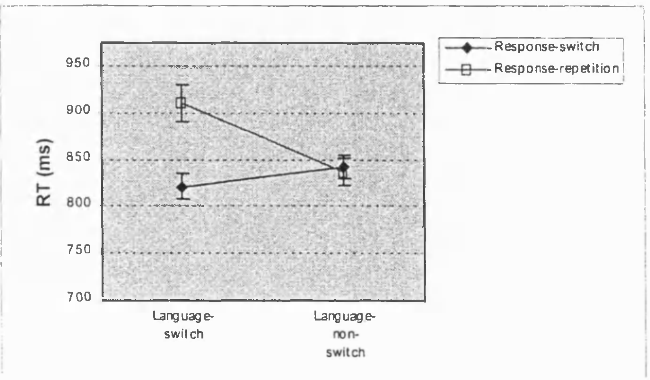
There was only one significant three-way interaction, namely between the factors part, language-switching and response type [$F_1(1,23) = 14.30$, MSE = 6809.67, $p = 0.001$; $F_2(1,126) = 7.25$, MSE = 11541.66, $p < 0.01$; min $F'(1,119) = 4.81$, $p < 0.05$]. This is interesting, as it indicates that whatever caused language-switching to interact with response type in the first half of the experiment, ceased doing so in the second half (see separate part 1 / part 2 analyses below). There were no further three-way interactions; critically, language-switching x response type was not modified by any other factor. There were no three-way interactions in the error analyses [Fs < 1 and *ps* > 0.1].

There were no four- or five-way interactions, in either the RT or in the error analyses (Fs <1 or *ps* > 0.1).

4.3.2.1 Language-switching and response type - in Part 1

Consistent with the prediction of the external account, there was a significant interaction between language-switching and response type [$F_1(1,23) = 56.05$, MSE = 3819.64, $p < 0.001$; $F_2(1,126) = 20.31$, MSE = 13080.09, $p < 0.001$; min $F'(1,136) = 14.91$, $p < 0.001$]. Please see Figure 4.1 and Table 4.2.

Figure 4.1
Part 1. Mean correct RT (ms) and standard errors of the mean, as a function of language-switching (language-switch / language-non-switch) and response type (response-switch / response-repetition).



As predicted by both the internal and external accounts, when repeating responses, RTs on language-switch trials were slower than RTs on language-non-switch trials [language-switching effect when repeating responses: $F_1(1,23) = 37.13$, MSE = 15833.98, $p < 0.001$; $F_2(1,126) = 23.33$, MSE = 16514.37, $p < 0.001$; min $F'(1,105) = 14.33$, $p < 0.001$]. A language-switch increased mean reaction time by 73ms [mean RTs of 910ms (196) and 837ms (134) for language-switch and language-non-switch respectively]. When changing responses, language-switch trials were *faster* than language-non-switch trials by 21ms [mean RTs of 821ms (133) and 842ms (131) for language-switch and language-non-switch, respectively; significant only by subjects: $F_1(1,23) = 18.32$, MSE = 4685.81, $p < 0.001$; $F_2(1,126) = 1.17$, $p > 0.2$; min $F'(1,140) = 1.10$, $p > 0.2$]. This pattern of a language-switch cost for one response type (response-repetition) but none for the other (response-switch) is explained with reference to language signals playing a cueing role via paralexical decision processes. Language-change signals lead to a bias towards a change of response - incorrect on response-repetition trials, engendering a delay to activating, sufficiently highly, the correct response units above the biased-towards incorrect ones.

In the case of errors, there was no reliable interaction between language-switching and response-switching [*p*s > 0.1 or Fs < 1]. What should be noted about the errors is that their pattern appears to go in the opposite direction to that of RTs on the response-repetition trials: while RTs decrease (language-non-switch trials), errors appear to become more numerous (see Table 4.2). It is possible that this is a case of speed/accuracy trade-off. However, if it were, one might expect to find such a trade-off on the other fast responses as well (language-switch/response-switch). In any event, the interaction is not reliably significant, and it may be that with a larger subject group the pattern would not be replicable. Above all, the error pattern does not invalidate the crucial RT finding, namely that responses were substantially delayed on response-repetition trials when language changed.

Inspection of Figure 4.1 lends support to the conjecture made in the Introduction of this chapter that a 'positive signal' (i.e., that something has changed [language-switch]) is stronger than a 'negative signal' (i.e., that there is no change [language-non-switch]): the RT delay on language-switch / response-repetition trials was greater than the RT delay on language-non-switch / response-switch trials.

Table 4.2

Part 1. Mean correct RT (ms; SD in brackets) and error percentage per cell, as a function of language-switching and response type (over languages and animacy type).

	Response Type (animacy-switching)			
Language-switching type	**Response-switch (animacy-switch)**		**Response-repetition (animacy-non-switch)**	
	Mean RT (*SD*)	Error %	Mean RT (*SD*)	Error %
Language-switch	821 (133)	6.8%	910 (196)	6.8%
Language-non-switch	842 (131)	8.2%	837 (134)	10.4%

4.3.2.2 Language-switching and response type - in Part 2

The data relevant to this section are to be found in Table 4.3. The language-switching by response type interaction that was found in part 1 was not obtained in part 2 [$F_1 < 1$; $F_2(1,126) = 1.05$, $p > 0.3$; min $F' < 1$]. In part 1, when repeating responses, there had been a large language-switch cost; in part 2, this effect was not found [$F_1(1,23) = 3.35$, MSE = 4475.40, $p = 0.080$; $F_2(1,126) = 2.19$, MSE = 15573.83, $p > 0.1$; min $F'(1,102) = 1.32$, $p > 0.2$]. Likewise, when changing responses, the effect of language-switching was not significant [$F_1(1,23) = 2.69$, MSE = 8709.44, $p > 0.1$; $F_2(1,126) = 1.54$, MSE = 12540.73, $p > 0.2$; min $F' < 1$]. As one may see in Figure 4.2, the large inhibitory effect previously seen on language-switch/response-repetition trials was no longer evident. In terms of errors, none of these interactions were significant [*p*s > 0.1 or F < 1].

This lack of a language-switch cost in part 2 is in accord with the predictions of the external account. With top-down inhibition from language-nodes to lexical representations, the internal account cannot accommodate the absence of language-switch costs.

Figure 4.2

Part 2. Mean correct RT (ms) and standard errors of the mean, as a function of language-switching (language-switch / language-non-switch) and response type (response-switch / response-repetition).

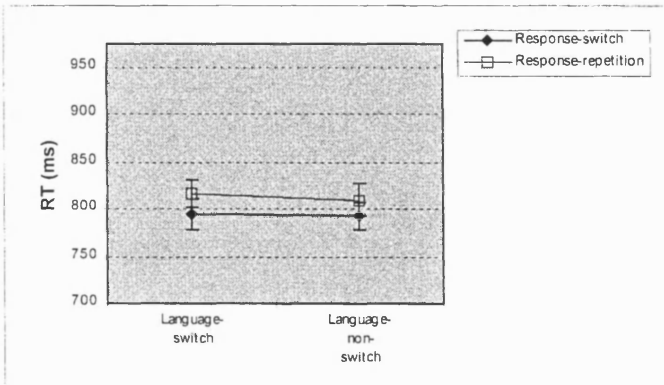


Table 4.3

Part 2. Mean correct RT (ms; SD in brackets) and error percentage per cell, as a function of language-switching and response type (over languages and animacy type).

	Response Type (animacy-switching)			
Language-switching type	Response-switch (animacy-switch)		Response-repetition (animacy-non-switch)	
	Mean RT (*SD*)	Error %	Mean RT (*SD*)	Error %
Language-switch	795 (153)	8.1%	817 (152)	8.1%
Language-non-switch	793 (136)	6.1%	810 (164)	7.3%

4.4 General Discussion

These data extend prior findings using a category-instance task (e.g., Dufour & Kroll, 1995), because the present findings suggest that the effects of language-switching may be modified by the nature of the response. Individuals responded more quickly on language-switch (than language-non-switch) trials when they were required to change their response from a prior trial, but they responded much more slowly to a switch in language when they had to make the same response (see Figure 4.1).

Critical to the external account is the idea that response times reflect the mapping of signals from the bilingual lexico-semantic system onto responses. It was supposed that the presentation of a word elicits a relatively automatic signal of its language membership (see Kroll & Dijkstra, 2001). The proposal here is that any change in the language of input, signals a change in the state of the world and this may cue a change in response, depending on how much weight individuals place on such language-change signals in the paralexical decision process[64]. It follows that a change in language will facilitate a response when participants do in fact have to change their response from the prior trial. In contrast, a change in language will hinder a reply when individuals have to maintain the same response. The data presented here indicate that the signal indicating that there has been a change in language interferes with generating the same response again. Both the internal and external accounts can accommodate the pattern of part 1.

Differences were found between parts 1 and 2 of the experiment with respect to language-switch costs (significant in part 1, n.s. in part 2, though the interaction part x language-switching was marginal by subjects and n.s. by items). The language-switching by response type interaction was significant in part 1 but n.s. in part 2. The comparison between Figures 4.1 and 4.2 shows that the large interference experienced on language-switch/response-repetition trials in part 1 disappeared in part 2. This outcome is open to a number of interpretations.

First, a switch in language may cease to be functionally relevant because individuals come to translate each word and so effectively activate both languages. However, Dufour & Kroll (1995) demonstrated in their bilingual category-instance task that individuals do not engage in spontaneous translation, even when given the opportunity to do so by presenting the instance well after the presentation of the category name (650ms delay).

A second possibility is that individuals entered a more bilingual mode (Grosjean, 1997, 2001) in the course of the experiment, because they processed words in both languages. Language changes could then be argued to cease carrying functional relevance because the lexical systems of both languages are equally engaged. However, the individuals participating in this experiment are likely to have been in a bilingual mode right from the start of the experiment: Bilinguals knew that they were taking part in a bilingual experiment and that the

---

[64] One might argue that any 'change in the world' could have such a cueing effect (e.g., changes of colour, font, size, location). To investigate whether language-switch signals are 'special', or whether any 'change in the world' can cue a response-switch (e.g., "yes" → "no"), one might contrast a condition crossing language-switching with response-switching, and a condition crossing (for instance) colour-switching and response-switching. If one found an effect only in the former condition, this would support the view held here, that language-switches are meaningful, important events (unlike other, random, changes) and can influence response selection. (See also Thomas & Allport, 2000, for a discussion of the importance of language-switching on response type.)

experimenter was bilingual in the same languages as they[65]; additionally, instructions were given in both languages. Certainly after the practice block, not only as of the second half of the experiment, participants would have been expected to have entered a 'bilingual mode'. Indeed, Grosjean (2001) speculates that a single word in another language might be sufficient to shift individuals towards a relatively equal activation of both languages.

A third possibility is that the cueing effects of a language switch are reduced because of item repetition. The words presented in part 2 were translations of the full set of items presented in part 1. One might argue that there could be a 'lexically based' repetition effect (cf. de Groot & Nas, 1991). That is, processing the first presentation (e.g., DOG) raises the baseline activation level of the translation equivalent (HUND ['dog' in German]). It seems unlikely though, that there could have been lexically based cross-language repetition priming effects in this experiment, since such effects appear to be very transient (Forster & Davis, 1984; de Groot & Nas, 1991), and in the present experiment translations were never in adjacent blocks, let alone on adjacent trials. Another possible cause for repetition effects is 'episodically based' (cf. de Groot & Nas, 1991). That is, participants might speed their responses by retrieving the decision they made about the item in part 1 (e.g., the meaning representation {dog} was categorised as 'animate' once before). One possible objection to this account is that error rates were comparable in parts 1 and 2. In any event, it is not clear why repetition priming (of whatever nature) should affect particularly one trial type: language-switch/response-repetition trials evoked much slower RTs than other trial types in part 1; but in part 2, these slow RTs were no longer apparent (contrast Figures 4.1 and 4.2). Furthermore, other literature suggests that item repetition is not obviously the reason for the part1/part 2 switch-cost difference found here. Shafiullah & Monsell (1999) suggest "that switch costs are resistant to amelioration by repetition." (p. 597). In an animacy decision task, in which participants switched between seeing stimuli in Kanji and Kana, they found similar switch-cost patterns in experiments repeating items vs. showing items only once. Specifically, there was a reduction of the temporal switch-cost over the course of the experiment, both in their no-repeated-items experiment (Experiment 1a) and in one of their repeated-items experiments (Experiment 1b). As in the present experiment, there was no switch-cost reduction for accuracy. (But note that script-switching and response type seem not to have

---

[65] Grosjean (1997, p. 229; 1998) points out these factors, amongst others, as contributing to shifting bilinguals towards the bilingual end of the unilingual/bilingual continuum.

been crossed, which the present Experiment 3 has shown to be important.) Undoubtedly, however, it would be valuable to assess the question of an adaptation to the cueing role of language signals using novel items throughout, so as to clearly rule out any repetition effects.

A fourth possibility, and the one favoured here, is that individuals come to reduce the weight attached to the language cue in the paralexical decision process. Under normal circumstances, a change of language is an event with consequences for language processing (e.g., different syntactic procedures may be required, depending on the input language - cf. Grosjean, 1997). Thus, reflecting habit, individuals may initially place much weight on language signals in their paralexical decision process. However, on half the trials, language cues are misleading, biasing towards the incorrect response. To avoid erroneous responses, individuals need to revoke that bias, and ensure that the other response is given. To avoid this problem of having to overcome a bias towards an incorrect response, individuals could reduce the weight placed on language signals, and hence reduce the problem of being erroneously biased towards the wrong response. That is, bilinguals adapt to the context of the task. An internal account with top-down inhibition (language-node level to lexical level), cannot accommodate such a lack of language-switch costs. It would only be able to predict the absence of this cost if it allows top-down inhibition of non-target lexical representations to be reduced to zero.

In sum, in part 1, before extensive experience (and before item repetition), the effect of a language-switch depends upon the nature of the response. The results presented here extend the findings of Thomas & Allport (2000) from the context of a lexical decision task to the context of a semantic categorisation task, where the language of input is manifestly irrelevant to the task and where the critical information is of a semantic rather than of a lexical nature. The results, specifically of part 2, are difficult to explain with an internal account of the costs of language-switching (i.e., as being due to activation decrements). That is, this experiment shows more clearly than Experiments 1a/1b and 2 were able to show, evidence supporting the external account but refuting an internal account that achieves control by inhibiting non-target lexical representations. Note that the input-switch and associative connections accounts would also have predicted clear language-switch costs, for all response types and for both parts.

It could be argued that a language signal (and hence language-change information) is available earlier than animacy information in the case of poor exemplars, which appeared to take longer to process, while in the case of good exemplars the meaning information is available more quickly, and so animacy information does not lag behind the language signal as much. However, consider that in the item analysis 50% of the items were good and 50% were poor exemplars. If language-change signals were only able to have the response biasing effect in the case of poor items, then the item analysis would not have shown language-switching to delay decisions on response-repetition trials.

The error analysis did not show increased errors on language-switch / response-repetition trials, which were clearly affected in terms of RTs. This suggests that individuals did not allow language-change signals to drive the response (nor that they simply acted along the lines of "is the current item the same or different from the previous one?"). Had either of these been the case, then the error rate should have increased, but RT would not have, on language-switch / response-repetition trials. The results indicate that the role of language was a cueing one, but not a deciding factor in the decisions. Furthermore, the data suggest that individuals can adapt to signals from their language system, but it cannot provide unambiguous evidence for this interpretation. Chapter 5 will revisit this question of adaptation.

As mentioned in the Introduction to this chapter, there may be both 'early' and 'late' language signals. Of the stimuli used in this experiment, about one third were post-experimentally judged, by the experimenter, to be illegal or at least unusual in the other language (e.g., SHIRT, LÄUSE, illegal for German and English respectively). To investigate whether the effect of early language signals differs from that of late language signals (e.g., to create a stronger bias, and hence (1) a greater delay on language-switch / response-repetition trials and (2) a greater benefit on language-switch / response-switch trials), one could contrast two different conditions. In one, stimuli would all be of the 'illegal in the other language' variety, in the other condition stimuli would all be legal in both languages. The external account would be able to accommodate such an effect of stimulus type, as paralexical decision processes play an important role in this account. An internal account with top-down inhibition (language-node to lexical level) would be able to accommodate greater delay on language-switch / response-repetition trials for the 'early signals' condition rather than 'late signals' condition, since an erroneous bias could be occurring while suppression of lexical representations was being overcome. It could also accommodate a decreased - possibly eliminated - language-switch cost for response-switch trials in an 'early signals' condition. However, in the 'late signals' condtion a language-switch cost would be expected both for response-switch and response-repetition trials, since in that case the biasing language signals only become available once the activation decrement has been overcome. At that stage it is too late for biasing to have an effect on RT.

4.5 Conclusion

The data presented here suggest that, although language-switching took place, the influence of language was not a simple language-switch cost. Instead, language-switching and response type interacted, suggesting that language-switch costs are influenced by the way in which bilinguals are asked to use their languages, which influences the nature of the task schemata. This point has so far been tested using lexical decisions (Experiments 1a/1b and 2). One might attempt to test this notion further by modifying the nature of the instructions in the animacy decision task. An animacy decision task does not lend itself as easily to language-specific decisions as a lexical decision task (where individuals can be asked to "decide whether the letter string is a word in a specific language"). However, one might attempt to create a language-specific task by telling participants that two of four squares (cf. Experiment 2) will always contain English words, and the two others will always contain German words. This way, participants should become fully aware of language-switches and be able to predict them. This might induce them to differentiate between languages, and perhaps even set up separate task schemata. In that case, a larger language-switch cost would be expected compared to the one found here. There should also be a stronger language-switching x response-type interaction. Furthermore, it is possible that in that case this pattern would be visible throughout, rather than disappearing, as was seen here in part 2.

The lack of a language-switch cost and of an interaction between language-switching and response type in part 2 could reflect that individuals reduced the weight placed on language signals, i.e., that individuals adapted their processing. This pattern also suggests that only a single task schema was employed, since with two schemata the external account would have predicted some remaining switch-cost. This would need to be tested further in a task using novel items throughout, avoiding the possible confound posed by item repetition in this experiment's part 2[66].

4.6 Outlook to Experiments 4a & 4b

In Experiments 1a/1b and 2 the effects of the non-target language on the target language could not be clearly shown with the nonword type effect. It was argued that this might be related to the nature of the materials, in that people have no representation of nonwords in their lexica (other than real-word neighbours). Experiments 4a/4b pick up this issue of cross-language interaction by investigating interlingual homograph (IH) interference effects. Secondly, Experiment 1b, the language-general task in Experiment 2 and the present experiment, looked at the effect of language when it is not relevant to the task, though overtly presented (such that there are language-switches). Language-switch costs became very small in these cases. In the present experiment it was even possible to provide an indication that language-switch costs need not be present even when switching languages (part 2). The small cost of language-switching in the present experiment has here been attributed to the effects of language-change signals on the units underlying responses. In the experiments presented so far, both languages were overtly presented, and so language information was obviously present. Experiments 4a/4b take the investigation of the effects of language signals further by asking what effect a language can have that is not only irrelevant to task demands (as in a language-general task), but is furthermore not even overtly presented. Thirdly, if it was possible to show adaptation to misleading language signals (part 2), will it be possible to find evidence of adaptation to the interference effects of an interlingual homograph's alternative reading? To address these questions, Experiments 4a/4b employ English-specific lexical decision tasks, which include some (non-cognate) interlingual homographs (words that have the same spelling in both languages, but different meanings, e.g., 'MUTTER' ['mother' in German]).

---

[66] To ascertain whether in a lexical task (an LDT), there are two task schemata (as suggested by Experiments 1a/1b and 2), an LDT could be carried out which likewise contrasts performance early and late in the experiment, and in which language-switching and response-switching are fully crossed. If it was found that language-switch costs existed only initially, but then disappeared, this would suggest that it is possible to have a single task schema even when the language-coded lexical representations are targeted. Otherwise, one might conclude (as hypothesised so far) that by force of habitually differentiating between languages at the lexical level, two task schemata are set up, one for each language.

V

Experiments 4a & 4b: The locus of controlling interlingual homograph interference

5.1 Introduction

In the preceding three experiments the bilingual participants' two languages were overtly present: item lists included both German and English words. The data from these experiments provide evidence about the origin of language-switch costs (Experiments 1-3), and they suggest that it is possible to adapt to the interfering influences of language signals (Experiment 3). However, the question of adaptation needs to be assessed more directly. Furthermore, so far there is no clear support for the non-selective access view which the external account takes: As discussed in Chapters 2 and 3, nonword type effects as investigated here, did not provide evidence for non-selective access. Experiments 4a and 4b, presented here, address both these issues. Critically, Experiments 4a and 4b contrast the internal and external accounts of control using a different paradigm to the preceding, to ensure that results obtained so far in favour of the external account are not task-specific.

The task employed here is an English-specific LDT (i.e., all decisions are whether or not the letter string is a word in English). Thus, there is no overt language-switching. The only relevant language is English. German is introduced in the form of German/English (non-cognate) interlingual homographs (identical spelling, but different meaning in the two languages, e.g., 'MUTTER'; 'mother' in German). If the two language systems were separate and independent, then RT on these critical items should show no effect of their dual-language status. Furthermore, as various experimental manipulations are designed to investigate, it is of interest to know whether it is possible to control cross-language interference on such items (assuming that the two language systems are interdependent), and what the locus of this control is. The various interlingual homograph (IH) experiments in the literature (e.g., Dijkstra, van Jaarsveld & ten Brinke, 1998; Dijkstra, de Bruijn, Schriefers & ten Brinke, 2000; de Groot, Delmaar & Lupker, 2000) have not fully addressed these questions. Further details are discussed in the following.

Non-cognate interlingual homographs (IHs, for short) offer a means of exploring the extent and locus of control processes. All German/English IHs employed here have a relatively low frequency of occurrence in English but a high frequency of occurrence in German. Given non-selective access, bilinguals performing an English-specific LDT should generally be slower to respond to IHs than to (purely English) control words matched to the IH's English frequency. IH interference should arise, at least in part, because the higher frequency reading is in German, and this will activate the units underlying the overt 'no' response; yet a 'yes' response is required, because the item *is* an English word. Given that 'yes' response units have to exceed the activation of competing 'no' units by a criterial amount, RT to an interlingual homograph will be delayed relative to that of its matched control word because the control word only has an English reading, thus giving rise only to signals activating 'yes' units (see Levelt, Roelofs & Meyer, 1999, for one competition-sensitive mechanism).

As mentioned above, previous research in the literature (e.g., Dijkstra, van Jaarsveld & ten Brinke, 1998; Dijkstra, de Bruijn, Schriefers & ten Brinke 2000; de Groot, Delmaar & Lupker 2000) has left open questions about the extent and locus of control.

Dijkstra, van Jaarsveld & ten Brinke (1998), using Dutch/English bilinguals, examined RT to IHs relative to controls in an English-specific lexical decision task, both in the absence of pure Dutch words (Experiment 1) and in their presence (Experiment 2). The pure Dutch words required a 'no' response. Surprisingly, and contrary to expectation, given the interference process described above, participants responded no more slowly to IHs than to matched control words when there were no pure Dutch words (Experiment 1). In contrast, when pure Dutch words *were* shown (Experiment 2), Dijkstra, van Jaarsveld & ten Brinke (1998) obtained strong IH interference (see also Dijkstra, de Bruijn, Schriefers & ten Brinke, 2000). Dijkstra, van Jaarsveld & ten Brinke (1998) proposed an *internal* account of their results (p. 58). In the absence of pure Dutch words, a more active English system is able to suppress activity in the Dutch system. In the presence of pure Dutch words, activation of the Dutch system is boosted bottom-up (see also Dijkstra & van Heuven, 1998), making suppression difficult, if not impossible. In this case, then, there is IH interference (slower RT for IHs than for matched controls)[67].

De Groot, Delmaar & Lupker (2000) explain these findings differently. Some individuals always (or all individuals some of the time) *reconstrued* the task in Dijkstra, van Jaarsveld & ten Brinke's (1998) Experiment 1. Participants acted as if the task was to respond 'yes' if the letter string is a word, whether in English or in Dutch (i.e., they reconstrued a

---

[67] According to Grosjean (e.g., 2001), such non-target words 'shift' the bilingual further towards the bilingual end of the unilingual-bilingual continuum, making the other language more active, thus increasing the chance of interlingual interference.

language-specific LDT as a language-general LDT). In this way, they overcame the difficulty of making English-specific (L2) lexical decisions. A balanced mixture of language-specific responses (that should yield IH interference) and language-general responses (that should yield IH facilitation when the Dutch reading is of higher frequency), could, fortuitously, account for the null-result. In the presence of pure L1 words (Dijkstra, van Jaarsveld & ten Brinke, 1998, Experiment 2; de Groot, Delmaar & Lupker, 2000, Experiment 3), de Groot, Delmaar & Lupker (2000) argue, reconstrual is blocked (see also Dijkstra, van Jaarsveld & ten Brinke 1998, footnote 4), and so IH interference becomes evident. For Dutch (L1) lexical decisions, de Groot, Delmaar & Lupker (2000, Experiment 2) found IH interference, though no pure non-target language words were presented. When making decisions in L1, individuals presumably had less reason to reconstrue the task.

In sum, this convenient means for examining issues of control in receptive processing, yields data open to very different accounts of control. There is a need to explore the nature and locus of control. Methodologically, there are two related comments on the studies just discussed. Each leads to a control question that shall be addressed by Experiments 4a and 4b respectively.

(1) It is not yet clear whether bilinguals are able to *modulate* IH interference. The IH interference effect reported in the presence of pure L1 words reflects performance following practice. In both de Groot, Delmaar & Lupker (2000, Experiment 3) and Dijkstra, van Jaarsveld & ten Brinke (1998, Experiment 2 - Dijkstra, personal communication, March, 1999) pure L1 words were presented in the practice list. The possibility is open, then, that individuals *can* reduce IH interference even when pure, non-target language words are included. In the experiments of Dijkstra, van Jaarsveld & ten Brinke. (1998) and de Groot, Delmaar & Lupker (2000), such modulation may have occurred *before* the experimental trials began. In the case of Dijkstra, de Bruijn, Schriefers & ten Brinke (2000), where pure L1 words were shown only in the second block of trials, there was no analysis of RTs to IHs and controls within that part; thus it is not known whether adaptation occurred. In sum, the studies by Dijkstra, van Jaarsveld & ten Brinke (1998), Dijkstra, de Bruijn, Schriefers & ten Brinke, (2000), as well as those by de Groot, Delmaar & Lupker (2000), establish that pure non-target language items increase IH interference. But they do not reveal whether such interference can be modulated, and hence do not address the question of the locus of such modulation. Experiment 4a examines these questions.

(2) The previous experiments on IHs (Dijkstra, van Jaarsfeld & ten Brinke, 1998; Dijkstra, de Bruijn, Schriefers & ten Brinke, 2000; de Groot, Delmaar & Lupker 2000 - referred to as 'the three Dutch studies' below) contrasted conditions containing pure non-target language words with conditions that excluded them. In the three Dutch studies, when pure non-target language words were included, individuals were informed about the presence of IHs. In conditions where they were not included, individuals were not informed that such items were in the list. If being informed modifies responses, then the comparison between 'with non-target words' / 'without non-target words' is tarnished. It is clearly critical to inform participants when non-target words are presented. If this is not done, then individuals will either be uncertain as to how to respond to interlingual homographs (in which case it will need to be explained to them and so they will become informed) or they will reject them (in which case there will be no RT data for correct responses). For these reasons, it was deemed unwise, when using stimuli such as MUTTER, to run a condition (in the current paradigm) that includes pure non-target language words without informing individuals about the presence of IHs. In the absence of pure non-target words though, it is possible to examine the effects of instructions. This is undertaken in Experiment 4b. Theoretically, being informed could allow individuals to reduce activation of the non-target system, e.g., by using the target language's language-node to inhibit non-target language representations (internal account). Alternatively, instructions could allow individuals to adjust the way they respond to signals coming from the bilingual lexico-semantic system (external account). For instance, instructions could influence the type of task schema individuals set up.

To summarise, the studies by Dijkstra, van Jaarsveld & ten Brinke (1998), Dijkstra, de Bruijn, Schriefers & ten Brinke (2000) and de Groot, Delmaar & Lupker (2000) establish that the presence of pure non-target language words increases IH interference, but these studies do not address whether such interference can be modulated, or where the locus of such control could be. Experiment 4a examines these questions. Prior studies also leave open the possibility that bilinguals, performing an English-specific LDT, and informed of the presence of IHs, can reduce the size of the IH interference effect when there are no 'pure' non-target language words (e.g., 'Seife', 'soap' in German) in the stimulus list. Experiment 4b examines this question.

Before commencing with the details of the experimental work, another theoretical issue needs to be addressed. Is a delay on an IH really due to competition between the yes/no response units (as suggested earlier), or could it simply reflect that individuals notice IH items and hesitate? That is, the response criterion could become more cautious, leading to delay. This hesitance could also be carried over to following item(s), causing delay on them. (Note, such a 'carry-over effect' can also be explained on the yes/no response competition account - see page 181/182.) This alternative explanation shall be examined further with recourse to the data pattern found here.

Experiment 4a:
Control in the presence of pure non-target language words

The performance of bilinguals in two conditions was contrasted. In both conditions participants were informed that they might encounter words that were words in both English and German and that they were to respond 'yes' to such words because they were English words. One condition was shown pure German words (the Informed+German[68] condition), whereas the other was not (the Informed-German[69] condition). The experimental trials were divided into two halves (experimental phases 1 and 2), allowing the experimenter to investigate adaptation. It is important to note that the practice list contained only pure English words and English-like nonwords (i.e., no German words, no IHs and no pseudohomophones of German words).

Despite the findings by Dijkstra, van Jaarsveld & ten Brinke (1998), Dijkstra, de Bruijn, Schriefers & ten Brinke (2000) and de Groot, Delmaar & Lupker (2000), it was expected that individuals participating in the present experiments would show IH interference even in the absence of pure non-target language words. Of course, participants in the Informed-German condition *might* elect to make language-general lexical decisions despite language-specific instructions, but there were prior grounds for believing that they would not reconstrue their task in this way: The language switching studies of this thesis (Experiments 1a/1b and 2), showed that the cost of switching between languages was significantly greater when individuals were asked to make language-specific rather than language-general lexical decisions. This increased switch-cost arose even though no 'wrong language' words were presented (e.g., German words on trials designated as English, and vice versa; cf. Thomas & Allport, 2000), which could have *enforced* language-specific responses. If bilinguals

---

[68] read as "Informed plus German"

[69] read as "Informed minus German"

confronted with a difficult task in their L2 reconstrue a language-specific LDT as a language-general one in the absence of non-target language words, then there should have been little or no difference in switch-costs between these tasks. There was, however, a clear difference between the two LDT types. This difference was considered indicative of participants adhering to task instructions. (This is particularly clear in the pure-HF condition in Experiment 2, where participants carry out the language-specific task after the language-general one: even in that case the switch-cost in the language-specific task was significantly larger than in the language-general task.) In short, IH interference was expected on the basis that participants would not reconstrue the task. The supposition as to how this IH interference would arise is as follows. Given non-selective access, IHs will give rise to two different language signals. As described above, 'English word' signals will be mapped onto the units for a 'yes' response, the earlier occurring 'German word' signal will map onto the units for a 'no' response. This response competition should yield IH interference.

However, in line with the data of Dijkstra, van Jaarsveld & ten Brinke (1998), Dijkstra, de Bruijn, Schriefers & ten Brinke (2000) and de Groot, Delmaar & Lupker (2000), it was expected that IH interference would be greater overall in the Informed+German condition. The experimental question was whether or not individuals in the Informed+German condition would be able to reduce the degree of interference. If control is possible (whether at an internal locus or at an external locus), then IH interference should be less in experimental phase 2 compared to experimental phase 1 (i.e., there should be a clear difference in the degree of IH interference manifested by the two conditions in phase 1, but less, or even none, in phase 2).

Assuming that adaptation occurs, its locus was explored by also examining carry-over effects. RTs to high frequency (HF) English words were recorded. Two were placed immediately following each IH, and two following its matched control word. Characteristics of previous stimuli are known to modulate RTs to immediately following stimuli. For instance, Lima & Huntsman (1997, Experiment 1), in a unilingual lexical decision task, showed that both word and nonword responses were slower when the previous stimulus was a nonword than when it was a word. In the present case, it was supposed that an IH, in contrast to its matched control word, produces greater residual activation in the 'no' response units. This response competition should delay the time required to reach a 'yes' response to the following HF word as opposed to a 'yes' response following a control word. Certain other assumptions shall be identified later in this chapter's General Discussion. The RT difference between HF words following IHs and HF words following controls, is here referred to as the 'carry-over effect'. It was expected that the carry-over effect would dissipate by the second high frequency word.

If it is the case that pure German words increase the activation of the German system (or the units coding for language membership) and lead, at least initially, to greater IH interference, then the carry-over effect should also be greater in the Informed+German condition than in the Informed-German condition. This is because greater activation of the German system will presumably entail greater activation of the 'no' units. With greater activation of 'no', it is more difficult to raise the activation level of the 'yes' units sufficiently in order to be able to respond 'yes' on the HF word that follows the IH.

The importance of the carry-over effects lies in that they make it possible, in combination with IH interference effects, to distinguish the two loci of control (external / internal), which the IH interference effect alone cannot achieve: If individuals are able to reduce the activation level of the German system (internal account), then not only will IH interference decrease (because there is less activation in the units underlying a 'no' response), there will also be less residual activation in these units to delay a 'yes' response on the following HF English word trial. In short, an internal locus of control predicts that a decrease in IH interference will co-occur with a decrease in the size of the carry-over effect. In contrast, an external locus of control predicts that IH interference and carry-over effects will dissociate. According to the external account, individuals reduce IH interference by changing how they respond to IHs. Thus residual activation in the 'no' units remains unchanged and so the carry-over effect will remain unchanged too. In short, the extent to which IH interference and carry-over effects co-vary or dissociate makes it possible to distinguish between the internal and external accounts of control. It is presumed here that the size of the IH interference and carry-over effects depends on the rate at which activation accumulates and decays in the response units. This could be subject to individual variation, but the predictions made here are independent of it since they rely, in the case of IH interference, on the difference in RT to IHs and their matched controls and, in the case of carry-over effects, on the differences in RTs to words immediately following them. In both cases the key measures are within-subject.

5.2 Method

5.2.1 Participants

32 German/English bilinguals (students at University College London and young professionals working in London), with a mean age of 29.5 years (*SD* = 5.5, range = 20-40), participated. They were randomly assigned to the two conditions, such that there were equal numbers of males and females in each. On average, bilingual participants had been in the UK for the past 7.4 (4.0) years. None had previous experience with the lexical decision task. 16 English unilingual participants (8 male, 8 female) with a mean age of 29.9 years (*SD* = 5.8, range = 20-40), provided a baseline measure of performance. All participants had normal or corrected vision. Participation was voluntary and was remunerated at the standard departmental rate.

Language proficiency:

Bilingual participants grew up with German and started to learn English at school as their second language at an average age of 9.1 years (2.4). Word recognition scores on the LLEX test (Meara, 1994) averaged 92.8 (3.4) out of 100 (range 83-99) for English, and 92.1 (4.4), range 83-99, for German, indicating comparable proficiency in the two languages on this test. For English unilingual speakers, the LLEX (English) test mean score was 94.8 (2.1), range 89-97. This shows that bilinguals in this experiment (as, indeed, bilinguals in the preceding experiments as well), had a high level of proficiency in their L2.

In the language-background questionnaire (see Appendix 1) these participants estimated that they used English more commonly than German [72.4 (20.0) out of a combined total of 100], but considered themselves somewhat more proficient in German [51.8 (7.4) out of a combined total of 100]. Bilinguals also rated themselves on a number of specific language skills (e.g., grammatical errors made, ability to adapt complexity levels to interlocutors and situations, ability to produce vocabulary and sentences at natives' standard; adapted from Bachman & Palmer, 1989; see Appendix 2). Their mean scores (out a maximum of 28) were 25.9 (1.6) and 24.3 (2.3) for German and English, respectively. English unilinguals rated themselves 24.6 (1.3) in English.

5.2.2 Design

The design for examining IH interference effects and that for examining carry-over effects are described separately.

a) Bilingual conditions:

IH interference effects. These effects were analysed using data from the IH and control word trials. Condition (Informed+German/Informed-German) was a between-subjects factor; Item Type (IH / Control) and experimental Phase (phase 1 / 2) were within-subjects factors. After the practice block, individuals in both bilingual conditions were informed about the presence of IHs and shown an exemplar.

Carry-over effects. These effects were analysed using data from the HF English word trials that immediately followed each IH and each control (see Table 5.1). In addition to the between-subjects factor Condition and the within-subjects factor Phase, there were two further within-subjects factors: Prior Item Type (following an IH / following a control), and Position (position 1 / 2), corresponding to the first and second high frequency words (referred to as HF1 and HF2 below) following each IH and its matched control. High frequency words were used in these positions to maximise the number of correct 'yes' reaction times, thus ensuring sufficient analytical power for the carry-over analyses.

b) Unilingual baseline condition:

For the IH/control investigation the within-subjects factors were Item Type (IH / Control) and experimental Phase (phase 1 / 2). Analysis of the carry-over effect had the same within-subjects factors as described for the bilinguals above.

Trial sequences:

The experiment comprised a practice block (consisting of 20 purely English words - i.e., no IHs - and 20 English-like nonwords) followed by the two experimental phases (phase 1, phase 2). There was an equal number of 'yes' and 'no' responses. Successive items were unrelated semantically or associatively and there was no stimulus repetition. Each experimental phase consisted of two blocks of trials (allowing a break half-way through each phase and between phases).

Each block began with 8 filler trials (4 real English words and 4 English nonwords in random order) followed by five 'critical sequences' and five 'filler sequences' in alternation. Each critical sequence contained, among other items, one IH and one control (further details in the next paragraph). The filler sequences each comprised four low frequency English words and six nonwords in pseudo-random order, such that word to nonword sequences were as frequent as nonword to word sequences. The filler sequences served as buffers before the next critical sequence. Thus ten IH/control pairs were presented per experimental phase.

Each critical sequence comprised 10 items. The items in this critical sequence were, in succession: 2 'no' trials, an IH, 2 HF English words, 2 'no' trials, a control word, and 2 HF English words. Depending on the condition, the two 'no' trials at the start of the critical sequence, and those in the middle, were either two English nonwords (Informed-German condition) or a pure German word followed by an English nonword (Informed+German condition). IHs therefore never occurred on trials involving a language shift (cf. Experiments 1a/1b and 2). For the Informed+German condition item sequences were identical in phases 1 and 2. For the Informed-German condition and the English uniliguals, phase 1 was as described above. In phase 2 a slight change took place, in that the first of the two 'no' item (an English nonword in phase 1) became a German nonword - all else was as in phase 1 (see Table 5.1). German nonwords were included in phase 2 to see if these alone would be sufficient to increase activation in the German system and thus IH interference (as well as carry-over).

Table 5.1
Experiment 4a. The sequential order of the trials in the critical sequences for each condition/group, and experimental phase.

					Critical sequence trials in experimental phase 1					
Informed+G	G_W	E_{NW}	IH	E_{HF1}	E_{HF2}	G_W	E_{NW}	Control	E_{HF1}	E_{HF2}
Informed-G	E_{NW}	E_{NW}	IH	E_{HF1}	E_{HF2}	E_{NW}	E_{NW}	Control	E_{HF1}	E_{HF2}
Unilinguals	E_{NW}	E_{NW}	IH	E_{HF1}	E_{HF2}	E_{NW}	E_{NW}	Control	E_{HF1}	E_{HF2}
					Critical sequence trials in experimental phase 2					
Informed+G	G_W	E_{NW}	IH	E_{HF1}	E_{HF2}	G_W	E_{NW}	Control	E_{HF1}	E_{HF2}
Informed-G	G_{NW}	E_{NW}	IH	E_{HF1}	E_{HF2}	G_{NW}	E_{NW}	Control	E_{HF1}	E_{HF2}
Unilinguals	G_{NW}	E_{NW}	IH	E_{HF1}	E_{HF2}	G_{NW}	E_{NW}	Control	E_{HF1}	E_{HF2}

G_W = 'pure' German word, G_{NW} = German nonword, E_{NW} = English nonword
IH = Interlingual Homograph, Control = matched control word (matched to English reading of IH)
E_{HF1} = first English HF word following the critical items (IH/control), E_{HF2} = second English HF word following the critical items (IH/control)

Within each critical sequence, IHs always appeared before controls. Should their order of appearance have been counterbalanced? Standard design procedure would dictate counterbalancing. For instance, it could be argued that if participants noticed IHs, this could cause an increase in caution (raising the response criterion level). The response criterion level could remain raised for a while, entailing artificially slowed RTs for following items, among them the control words. In short, it could be argued that the control words would provide a better baseline measure for IH RTs if they were presented before IHs, or rather, to avoid any order effects, that the presentation order of IHs and controls should be counterbalanced. However, the role of control words here was not to provide an entirely neutral response baseline. In this experiment, average (median) RTs were established for IHs and controls and the difference between them calculated. The differences (IH-control) arising in the different conditions were compared. In short, the focus of the experiment was on the *difference of differences* (i.e., how an IH interference effect varies as a function of condition). It was not the aim to establish a definitive size of an IH interference effect. Had the latter been the aim, then counterbalancing would indeed have been necessary. Thus, even if one believed there to be a confound (artificially raised RTs on control words), then this would be present for all conditions, and not affect the targeted difference of differences.

The data presented later on will, furthermore, indicate that while RT was delayed on IHs, and while this delay did carry over to the immediately following item (E_{HF1}), by the second item (E_{HF2}) the delay was gone (see section5.3.3). That is, it is unlikely that RTs on controls were affacted by the fact that IHs *preceded* them. If IHs were noticed and considered difficult, then RTs may have been slowed *overall*, not just for controls, but in the entire experiment.

If the delayed RTs on IHs were due to participants having noticed IHs, then presumably participants would be able to convey knowledge of them. As reported later on (end of section 5.2.4), at the end of the experiment participants were given a list of all IHs and controls (to assess knowledge of items); participants were surprised when the dual-language status of the IHs was shown to them. Participants reported not to have noticed this during the experiment.

In sum, while standard design procedure would have been to counterbalance the presentation order of IHs and controls, the data pattern investigated is not thought to have been compromised by the chosen design.

Item assignment:

The assignment of the IH/control words to phases 1 and 2 was rotated across participants within each condition and the order of the blocks was counterbalanced within each phase. The materials for assessing carry-over were held constant over these manipulations. That is, a given pair of HF words was always presented in the same order and remained coupled to the same IH or to the same control word, as these two types of stimuli were rotated and counterbalanced. Other solutions were considered impractical. For instance, the alternative of rotating a given HF word to each position after an IH, and to each position after its matched control, for the two experimental phases, requires a minimum of eight participants. For a meaningful item analysis (i.e., six to eight data points per cell), 48 to 64 participants would be required per condition. It would have been very difficult to procure adequate participants in such numbers.

5.2.3 Materials

a) Interlingual homographs and their matched control words:

From an initial rating study[70] (involving 20 unilingual English-speakers in London and 20 native German-speakers in Munich) 25 IHs (out of an initial sample of 104) were selected which had a higher subjective frequency of usage in German than in English (on a 7-point rating scale a minimum of three points difference on the average English and German ratings). The native German speakers had provided ratings for usage in German, the native English speakers had provided ratings for usage in English.

Several purely English words were then matched to each IH in terms of word length (number of letters and syllables), grammatical type, word frequency (Kučera & Francis, 1967), concreteness and imageability (derived from Appendix 2 of the MRC Psycholinguistic Database User Manual, Coltheart, 1981). Next, randomised lists of these 25 IHs, their potential controls, nonwords and additional English words (fillers) were presented to 40 German/English bilinguals (in Berlin, Bonn, London and Munich), asking them to mark all real English words. Words recognised as English words by a minimum of 70% of the raters were selected for the experimental set of IHs and controls, yielding 20 matched IH / control pairs (see Table 5.2).

Post-experimentally, an additional check of these words' frequencies was undertaken using CELEX (Baayen, Piepenbrock & van Rijn, 1993). CELEX (English) word frequencies correlated significantly with the Kučera & Francis (1967) frequencies that had been employed for the matching procedure (Pearson $r = .791, p < 0.001$). However, in the CELEX norms, the mean control words' frequency [13.4 (16.1)] was somewhat higher than the IHs' English reading, [8.9 (9.1)], though not significantly so [$F_2(1,19) = 2.17, p > 0.15$]. With respect to the frequency values of the IHs' German vs. English readings, the CELEX values confirmed

---

[70] None of the bilinguals taking part in any of the rating studies described here participated in the experiment proper.

that the mean value was significantly higher in German [118.85 (207.12)] than in English [8.9 (9.1)] [$F_2(1,19) = 5.55$, MSE = 21779, $p < 0.05$]. Table 5.2 shows the mean parameter values of the materials. Appendix 11 provides a full listing of the stimuli.

Table 5.2
Mean values (standard deviation in brackets) of word frequencies (CELEX; Kučera & Francis [K&F]; Ruoff), concreteness ratings (MRC Psycholinguistic Database), subjective frequency ratings (German; English [see text for details]) and word length (letters; syllables) for interlingual homographs (IH) and their controls.

	Matching parameters								
	CELEX frequency		**K & F frequ.**	**Ruoff frequ.**	**concreteness**	**subjective frequency**		**word length**	
	English	**German**	**English**	**German**	**English**	**English**	**German[1]**	**letters**	**syllables**
IH	8.9 (9.1)	118.9 (207.1)	13.3 (18.0)	136.3 (293.3)	530.3 (85.1)	2.96[1] (.62) 3.67[2] (.92)	4.67 (.56)	4.2 (1.0)	1.4 (0.5)
Control	13.4 (16.1)	N/A	13.4 (16.0)	N/A	548.5 (98.7)	4.25[1] (.74) 4.75[2] (1.11)	N/A	4.2 (1.0)	1.2 (0.4)

[1] given by German raters
[2] given by English raters

As a further check, also post-experimentally, on the *subjective* frequency of the chosen IHs and controls, randomised lists were constructed, comprising 120 stimuli - embedding the IHs and their controls amongst filler stimuli (HF and LF English words, pure German words, English nonwords). 24 English unilinguals indicated which items were nonwords and judged the frequency of usage of words existing in English. 24 German/English bilingual raters (with the same language background and ability as the ones participating in the RT experiment) performed the same task. They also performed a German version of it. Ratings were on a 7-point scale, with the end-values "used very frequently" and "used very infrequently" suitably rotated over participants. Of the bilingual raters, half were given the English version first, half the German version first.

Unexpectedly, given the Kučera & Francis (1967) ratings which had been used to match the stimuli (but in accord with the CELEX values), for the English unilinguals mean subjective frequency for the IHs was lower than for the controls: 3.67 (.92) vs. 4.75 (1.11)

[$F_1(1,23) = 65.88$, MSE = 0.21, $p < 0.001$; $F_2(1,19) = 33.45$, MSE = 0.47, $p < 0.001$; min $F'(1,36) = 22.19$, $p < 0.001$]. For the bilinguals (English ratings) a similar picture was found, 2.96 (.62) vs. 4.25 (.74) [$F_1(1,23) = 132.35$, MSE = 0.15, $p < 0.001$; $F_2(1,19) = 42.16$, MSE = 0.50, $p < 0.001$; min $F'(1,31) = 31.97$, $p < 0.001$]. See Table 5.2. The gap between subjective ratings for IHs vs. controls was slightly larger in the bilingual group, but a comparison of the ratings of these two groups showed that there was *no* significant interaction between item type and group [$F_1(1,46) = 1.43$, n.s.; $F_2(1,19) < 1$; min $F' < 1$], indicating that the difference in subjective frequency of IH items and controls was unaffected by language background. This is an important point which shall be picked up in this chapter's General Discussion. As expected, within the bilingual group, the IHs' German reading was rated as more frequent than their English reading: 4.67 (.56) vs. 2.96 (.62) [$F_1(1,23) = 147.01$, MSE = 0.24, $p < 0.001$; $F_2(1,19) = 45.19$, MSE =.68, $p < 0.001$; min $F'(1,30) = 34.56$, $p < 0.001$].

b) English words:

English words (for use in the critical sequences and as filler items) were chosen using the Kučera & Francis (1967) norms. They were orthographically legal in either language, though some were perhaps less usual for German (e.g., 'BRAIN'). None were interlingual homographs. The low frequency words ranged from 1-11 per million and the high frequency words ranged from 40-601 per million. Comparatively more low frequency than high frequency words (29% vs. 21%) were included in order to limit the distinctiveness of the IHs (LF in English) and to restrict response errors to IHs owing to short decision deadlines.

The HF words following the IHs and their controls matched the respective IH/control stimuli which they followed in terms of syllable and letter length. They were also matched in sets of four (two to follow IHs, two to follow controls) as closely as possible for frequency, concreteness and syllable/letter length. Thus these HF English words remained in sets of four, associated with the specific IH and control words which they matched for letter and syllable length.

c) Items for 'no' responses (nonwords and 'pure' German words):

English and German nonwords were constructed by exchanging a vowel in a real word not otherwise used in the experiment, see Method section of Chapter 2 for details. The nonwords were of the 'unique' type (i.e., legal only in the respective language on the basis of their consonant clusters; e.g., 'PFLOG' for German, 'SWERM' for English). The nonwords were matched, as far as possible, to the real words in terms of letter length and syllable length, though nonwords were on average half a letter longer than the real words (due to the consonant clusters).

50% of the stimuli were nonwords. For the Informed-German and unilingual conditions 44.2% of stimuli were English nonwords, 5.8% were German nonwords (German nonwords were shown as of phase 2). For the Informed+German condition 41.3% of stimuli were English nonwords, 8.7% were pure German words (pure German words were shown in both phases 1 and 2). The pure German words were high frequency [mean = 115 (164) in the Ruoff (1990) database and mean = 151 (246) in the CELEX (German) database]. Pure German words were nonwords in English (i.e., not pseudohomophones of real English words; cf. Nas, 1983).

'Englishness' of the English nonwords and German words:

Both pure German words and English nonwords needed to be "English-like" in these experiments. Individuals in the Informed+German condition were intended to identify pure German words as such (and not to reject them on orthographic grounds alone) and the English nonwords needed to be English-like especially for items immediately preceding the IH/control stimuli in order obviate any subjective language switch. A 7-point rating scale with the end values "couldn't possibly be an English word" and "could be an English word" (suitably rotated over participants) was used. Consistent with the requirements, 24 English unilinguals judged all stimuli to be acceptable for English, though the English-like nonwords (e.g., SWERM) had slightly higher means [5.00 (0.88)] compared to the pure German words (e.g., BRUDER) [4.54 (1.02)], $F_1(1,23) = 4.47$, MSE = 0.56, $p < 0.05$; $F_2(1,39) = 4.22$, MSE = 0.70, $p < 0.05$; min $F'(1,59) = 2.17$, $p = 0.14$ (cf. Altenberg & Cairns, 1983).

5.2.4 *Apparatus & Procedure*

Participants were contacted and instructed using English only. Each was tested individually in a small cubicle. Instructions and stimuli were presented on an IBM-compatible PC. Prior to the practice block, participants were instructed that on each trial a letter string would appear on the computer screen and that their task was to decide as quickly and accurately as possible whether or not it was a real word in English. Participants pressed a key of the computer keyboard to move through the initial instructions and to initiate each block of trials.

Stimuli were presented in white capital letters centred on a black screen. Responses were registered via two buttons (of a specially designed button box) marked '+' ("a real word in English") and '-' ("not a real word in English"). 'Reaction time' was the stimulus-response interval. All participants pressed the '+' button with the index finger of their dominant hand and the '-' button with the middle finger of the same hand. Each stimulus remained on the screen until a response had been given. The response-stimulus interval was 750ms. After the practice block (40 trials), and after each experimental block (108 trials), participants were told how many errors they had made and encouraged to keep this number under 10% while maintaining speed. There was no feedback regarding speed of responses.

Following the practice block, bilingual participants were told about the presence of IHs among the stimuli and how to respond to them ("In the experiment you may at times see words which exist in German. If such an item exists in German ONLY, reject it. If it exists in BOTH German and English, like 'KIND', accept it, because it *is* a word in English"). English unilinguals were throughout simply given English LD task instructions, no mention was made of German.

After the experiment, participants were presented with a randomised list comprising all IHs and control words. They marked the ones that they did not know as English words (in order to differentiate between errors due to lack of knowledge and processing errors). All bilinguals then carried out the English and German word recognition tests (LLEX, Meara, 1994), filled in the language background questionnaire and completed the adapted version of the Bachman & Palmer (1989) rating scales. Unilinguals completed the respective English versions only.

5.3 Results and Discussion of Experiment 4a

5.3.1 Data treatment

Given the restricted sample of IHs, the median was considered to offer a better measure of central tendency, though, in fact, analysis of the mean RTs, excluding outliers (RTs 2.5 *SD* from the treatment mean), showed the same pattern. RT analyses were on correct RT data. Error analyses were based on arcsine transformed error proportions. Over all bilingual participants, there were 5.3% errors on IHs and 0.6% errors on control words.

5.3.2 IH interference effects

The average correct lexical decision times for IHs and their matched control words are reported in Table 5.3. Average RTs in tables and text are mean of median ('M') and are from the subject data.

Table 5.3

Experiment 4a. Mean of median correct RT (ms), standard deviations (in brackets) and error percentages of IHs and controls, as a function of condition/group, experimental phase and item type.
(IH = interlingual homograph; G = pure German word).

	Phase 1				Phase 2			
	IH		Control		IH		Control	
Condition / group	**RT (*SD*)**	**Error %**	**RT (*SD*)**	**Error %**	**RT (*SD*)**	**Error %**	**RT (*SD*)**	**Error %**
Bilinguals: **Informed –G**	786 (106)	5.0*	724 (83)	1.3	770 (118)	1.9	692 (89)	0.6
Bilinguals: **Informed+G**	1170 (522)	6.9*	823 (328)	0.6	918 (339)	7.5	788 (261)	0.0
Unilinguals	632 (94)	5.0	605 (90)	1.3	622 (82)	4.4	583 (87)	1.3

* Includes 0.6% knowledge errors (i.e., errors on items not known as English words).

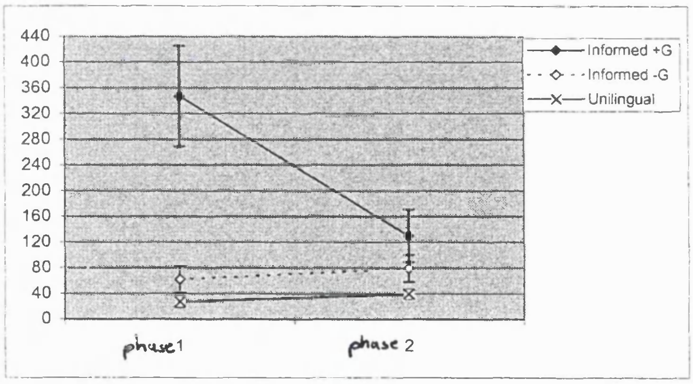
A mixed-factor ANOVA was carried out on these data. For the subject analyses Condition (Informed+German / Informed-German) was a between-subjects factor, while Item Type and Phase were within-subjects factors. In the item analyses Item Type was a between-subjects factor. The principal findings are reported here, Appendix 8 provides further details.

RT analyses: Analysis confirmed that bilinguals were slower to respond to IHs (M = 911ms, *SD* = 352) than to their matched control words (M = 757ms, *SD* = 219) [$F_1(1,30)$ = 29.18, MSE = 26090.02, $p < 0.001$; $F_2(1,76)$ = 33.32, MSE = 22140.87, $p < 0.001$; min F'(1,81) = 15.56, $p < 0.001$]. Further, as predicted by both the external and internal accounts, the size of the interference effect (i.e., the RT difference between IHs and controls) varied with condition (marginal on min F') [$F_1(1,30)$ = 8.69, MSE = 26090.02, $p < 0.01$; $F_2(1,76)$ = 5.67, MSE = 22140.87, $p < 0.025$; min F'(1,100) = 3.43, $p = 0.067$]. The overall interference effect was greater in the Informed+German condition (238ms) compared to the Informed-German condition (70ms).

Most critically, the size of the interference effect varied as a function of condition and experimental phase. The third-order interaction between item type, condition and phase was significant (marginal on min F') [$F_1(1,30)$ = 12.91, MSE = 8510.09, $p = 0.001$; $F_2(1,76)$ = 4.63, MSE = 9527.41, $p < 0.05$; min F'(1,106) = 3.41, $p = 0.068$]. In the Informed+German condition, the IH interference effect was 347ms in phase 1 and 130ms in phase 2. In contrast, in the Informed-German condition the comparable data were 62ms (phase 1) and 78ms (phase 2). The difference between phases in the size of the IH interference effect was significant in the Informed+German condition's subject analysis, and marginal by items and min F' [$F_1(1,15) = 14.11$, MSE = 13405.03, $p < 0.01$; $F_2(1,38) = 3.83$, MSE = 16130.59, $p = 0.058$; min $F'(1,52) = 3.01$, $p = 0.089$]. For the Informed-German condition the size of the IH interference effect was not modified by phase [all Fs < 1]. In short, 1) there was IH interference even in the absence of pure non-target language words; 2) even in the presence of pure non-target language words (Informed+German), bilinguals were able to reduce IH interference; 3) German nonwords (introduced in phase 2, Informed-German condition) did not increase IH interference. See Figure 5.1.

Figure 5.1

Experiment 4a. IH interference effect (ms), as a function of Condition (Informed+German / Informed-German) and Phase (1 / 2). (Analyses including the Unilingual baseline condition follow).



Error analyses: By requiring individuals to identify, post-experimentally, which IH/control stimuli they knew to be English words, it was possible to differentiate between errors due to a lack of knowledge and errors due to slips of processing. As can be seen (Table 5.3), errors were predominantly processing slips. Bilinguals made more errors on IHs than on controls [$F_1(1,30) = 23.60$, MSE = 0.01, $p < 0.001$, $F_2(1,76) = 17.45$, MSE = 0.02, $p < 0.001$; min $F'(1,96) = 10.03$, $p < 0.01$], and this was more pronounced in the Informed+German condition than in the Informed-German condition (a trend on min F') [$F_1(1,30) = 5.41$, MSE = 0.01, $p < 0.05$, $F_2(1,76) = 5.03$, MSE = 0.02, $p < 0.05$; min $F'(1,89) = 2.61$, $p = 0.11$] (see Table 5.3). In fact, although the third-order interaction between condition, phase and item type was not significant [Fs < 1], inspection of Table 5.3 suggests that the two-way interaction is largely the result of fewer errors on IHs in phase 2 for participants in the Informed-German condition. There is no immediately obvious reason for this reduction, but the crucial point to note is that a decrease in the size of the interference effect in the Informed+German condition did not co-occur with an increase in processing slips. That is, there is no evidence of a speed-accuracy trade-off as the basis of the effect.

Unilingual data: Unexpectedly (see Table 5.3), there was a small difference in RT between the IHs and their matched control words in the unilingual condition [$F_1(1,15) = 25.33$, MSE = 675.00, $p < 0.001$; $F_2(1,38) = 12.49$, MSE = 2511.28, $p < 0.01$; min $F'(1,52) = 8.37$, $p < 0.01$]. Note, however, that this difference was invariably greater for the bilingual participants: Comparing Informed+German with the unilinguals, the item type x condition interaction was significant ($p < 0.001$ - subjects and items, see Appendix 8 for details). Comparing Informed-German with the unilinguals, the item type x condition interaction was significant by subjects though not by items ($p < 0.05$ and $p = 0.14$; see Appendix 8 for details). Unilinguals also made more errors on IHs compared to their matched controls. In this case, though, there was little or no difference in error rates between unilingual and bilingual participants. In particular, the item type x condition interaction was not significant when contrasting the unilingual condition and the Informed-German condition [Fs < 1 or $p > 0.1$]. When contrasting the unilingual condition and the Informed+German condition, the item type x condition interaction showed a trend towards significance by subjects and items (n.s. on min F') [$F_1(1,30) = 2.89$, MSE = 0.01, $p = 0.10$; $F_2(1,76) = 2.59$, MSE = 0.02, $p = 0.11$; min $F'(1,89) = 1.39$, n.s.].

As described in the Method section, available frequency norms (Kučera & Francis, 1967) were used to match the IHs and their controls. However, the subsequent norming studies (see Method) showed that IHs had lower *subjective* frequencies than their controls. Conceivably, a difference in subjective frequency between IH and control items explains the differences in RT and error rate between the IHs and their controls in the unilingual condition. Given that bilinguals are slower than unilinguals, one might wish to argue that the IH interference effects are merely amplifications of a materials difference (whatever its source) shown in the unilingual data. Different lines of argument speak against this proposal. First, any materials difference between IHs and their controls cannot explain the observed change in IH interference as a function of condition and phase. The same IHs and their matched control words were presented in each condition, and in each condition materials were rotated over experimental phases[71]. Second, the carry-over effects, reported below, indicate a contribution of the German reading of the IH. There is another indication that IH interference is not simply a scaling effect (i.e., a materials difference that is amplified in the bilinguals). A correlation was undertaken to investigate whether the size of the IH interference effect increased with increasing RT to the control words. For none of the conditions or phases was the Pearson correlation between the degree of IH interference and control word RT significant (*p* was always at least greater than 0.2). That is, greater IH interference was not obviously a product of comparatively slower RTs in bilinguals than in unilinguals.

5.3.3 Carry-over effects

Above, evidence was obtained that IH interference can be modified in the presence of pure German words. Carry-over was analysed to determine the locus of control. To this end, RTs to the two high frequency English words (HF1 and HF2) directly following IHs and their controls were examined.

Carry-over effects were examined in a mixed-factor ANOVA with condition (Informed+German / Informed-German) as the between-subjects factor and prior item type (following IH / following control), position (HF1 / HF2) and experimental phase (phase 1 / phase 2) as the within-subjects factors. The RT analysis is reported for the subject data only. (This is not considered problematic, given the view that when item variability is experimentally controlled by matching with respect to variables that correlate highly with the dependent variable, a subject analysis is deemed sufficient - see Raaijmakers, Schrijnemakers & Gremmen, 1999.) As explained earlier (see Design section), a given pair of HF words was

---

[71] The materials difference was explored in more detail. The first analysis was a correlation of the rank order of the differences in median RT between each IH and its matched control word in each experimental phase in the unilingual condition with the rank order of these differences in the bilingual conditions (both Experiments 4a and 4b). In only one case (phase 2, Informed-German condition) was the rank order correlation significant ($p < 0.001$). Next, the possible basis for the materials difference was explored by correlating the differences between each IH and its control in subjective word frequency, in CELEX word frequency and in bigram frequency, with the RT differences. None of these factors proved robust predictors of the RT differences. IH and control words were also classified for (English) ambiguity of meaning. Words might be directly ambiguous (e.g., SAGE [herb/wise person] or indirectly ambiguous via a homophonous second meaning (e.g., HERD/'heard']. There were 6 IH/control pairs where only the IH was ambiguous; 3 pairs where only the control was ambiguous; 1 pair where both items were ambiguous and 10 pairs where neither word was ambiguous. The IH/control RT differences for the pairs falling into each category showed no obvious pattern. These explorations leave unclear what led to the unilinguals' IH/control RT difference, but whatever the source of the difference, it cannot explain the pattern of data in the bilingual conditions without taking account of the fact that an IH elicits a German reading.

always presented in the same order and remained coupled to the same IH or the same control word, precluding a comparable analysis by items. Errors on the items were minimal, also precluding analysis.

It was expected that, regardless of prior item type, RT to the second high frequency word (HF2) would be faster than to the first (HF1) because of response repetition (e.g., Rabbit, 1968; Pashler & Baylis, 1991; Thomas & Allport, 2000). This outcome was obtained in all conditions ($p < 0.05$), but shall not be discussed further because it is not directly germane to the issues addressed by the experiment.

To facilitate understanding, Table 5.4 presents the calculated carry-over effects (i.e., the difference in RT to an HF word following an IH compared to the RT to an HF word following a control word), rather than the median RTs for each HF word following an IH and following a control word.

Table 5.4
Experiment 4a. Carry-over effect[1] (based on mean of median correct RT, ms) and standard deviations (in brackets), as a function of condition/group, phase and position.
(G = pure German word).

	Phase 1		Phase 2	
Condition / group	**Position 1**	**Position 2**	**Position 1**	**Position 2**
Bilinguals: **Informed+G**	49 (82)	40 (61)	64 (67)	34 (75)
Bilinguals: **Informed-G**	21 (82)	10 (37)	46 (53)	22 (59)
Unilinguals	3 (37)	13 (54)	15 (62)	21 (63)

[1] Carry-over effect: calculated as [RT to high frequency word following IHs] minus [RT to HF word following controls], scored separately for the first HF word following IH/control (position 1) and the second HF word (position 2).

Errors were less than 0.8% in each condition, and did not change as a function of phase, previous item type or position; they are for this reason not included in Table 5.4.

Note: Table 5.4 only presents the calculated carry-over effects themselves to facilitate reading the table. Relevant mean RTs for HF words following IHs and those following controls are cited in the text.

Predictably, RTs were faster in phase 2 than in phase 1 [$F_1(1,30) = 10.34$, MSE = 12321.88, $p < 0.01$], phase 1: M = 705ms (245), phase 2: 661ms (183). More importantly, on the RT data, there was a significant effect of prior item type [$F_1(1,30) = 51.08$, MSE = 1612.51, $p < 0.001$]. RTs were slower after IHs (M = 701ms) than after their matched controls (M = 665ms). As predicted, there was also a significant interaction between prior item type and condition [$F_1(1,30) = 4.78$, MSE = 1612.51, $p < 0.05$]. The carry-over effect averaged 47ms in the Informed+German condition (M = 757ms following IHs and M = 710ms following controls) and 25ms in the Informed-German condition (M = 645ms following IHs and M = 620ms following controls). Such an outcome is consistent with the notion that pure German words increase the activation of the German system ("bottom-up") and so increase residual activation in the units underlying the 'no' response. As expected, the carry-over effect showed a tendency to decrease by HF2 [prior item type x position: $F_1(1,30) = 2.89$, MSE = 1819.98, $p = 0.10$] - see Table 5.4, Position 1 vs. Position 2.

Crucially, the three-way interaction between condition, prior item type and phase was *not* significant [$F_1 < 1$]. In short, the carry-over effect was no less in phase 2 than phase 1 for the Informed+German condition, *despite* a reduction in IH interference in phase 2 for that condition. This result is inconsistent with the idea that individuals reduced IH interference by reducing the activation of the German system (or the units coding for language membership). If they *had* reduced activation of the German system, then carry-over effects would have been reduced in tandem with IH interference, because there would have been less residual activation in the units underlying the 'no' response.

What of the notion that IH interference and carry-over are due to a criterion change, rather than response competition? The IH and carry-over data patterns of phases 1 and 2 speak against this notion. If a criterion change were responsible for the delay on IHs in phase 1 (as well as the carry-over effect there), then presumably the decreased IH interference (phase 2) would mean that, after some practice, individuals were less uncertain about responding to IHs: on seeing an IH their response criterion rose less than it had before. However, if IH interference and carry-over reflect a raised response criterion, then how can one explain that while IH interference decreased, carry-over remained unchanged? A response criterion change account cannot accommodate this divergent pattern. The pattern can be more easily explained with the notion of how signals are treated by the task schema, specifically, how they are linked to the units underlying the 'yes' and 'no' responses. This notion is explained in section 5.7.1.

If the carry-over effect reflects residual activation in the units underlying the 'no' response, occasioned by German reading of the IH, one should expect no carry-over effect for the unilinguals. As Table 5.4 indicates, there was no evidence of an effect of prior item type for the unilinguals (RT averaged 575ms following IHs and 562ms following controls - a difference that was not statistically significant [$F_1(1,15) = 2.43$, n.s.]).

Contrasting the carry-over effect in the Informed+German condition with that in the unilinguals, it was found, consistent with expectation, that prior item type and group interacted significantly [$F_1(1,30) = 9.94$, MSE = 1863.02, $p < 0.01$]. Although the overall third-order interaction (group x prior item type x position) was not significant [$F_1(1,30) = 1.55$, n.s.], the difference between the bilinguals (Informed+German) and the unilinguals was, as expected, significant on HF1 [$F_1(1,30) = 9.05$, MSE = 1984.01, $p < 0.01$]. It was not significant on HF2 [$F_1(1,30) = 1.98$, n.s.]. Carry-over effects were also contrasted between the Informed-German bilinguals and the unilinguals. Although the overall interaction between prior item type and group was not significant [$F_1(1,30) = 1.22$, n.s.], the interaction between prior item type and group showed a trend towards significance on HF1 [$F_1(1,30) = 3.03$, MSE = 1566.68, $p = 0.092$]. There was no effect by HF2 [$F_1 < 1$]. Presumably, for both bilingual conditions, the residual activation in the 'no' response units (occasioned by the German reading of the IH), had decreased sufficiently by the HF2, so that they no longer differed from the English unilinguals.

5.3.4 Responding to German words and English-like nonwords

In the Informed+German condition, individuals correctly rejected pure German words (e.g., 'BRUDER') and there was no difference in the error rates on pure German words and English-like nonwords (4.1% and 3.0 %, respectively). More important is the fact that error rates on pure German words were *unaffected* by phase [all Fs < 1] (4.1% and 3.8%, respectively). Clearly, individuals did not reduce IH interference in phase 2 by reconstruing their task as a language-general one. If they had done so, errors rates on pure German words would have increased in phase 2 compared to phase 1 (which did not happen). Further, individuals rejected these German words [M = 1111ms (413)] no more quickly than English-like nonwords [M = 1093ms (640)], Fs < 1. This outcome was expected both because these pure German words conformed orthographically to the rules of English (and so could not be rejected using an orthographic strategy, as one might for nonwords such as 'PFEDE') and because the presence of IHs ensured that individuals had to establish that these pure German words were not in fact English words.

5.4 Summary of Experiment 4a

Experiment 4a confirms prior results (Dijkstra, van Jaarsveld & ten Brinke, 1998; Dijkstra, de Bruijn, Schriefers & ten Brinke, 2000; de Groot, Delmaar & Lupker 2000), that presenting pure non-target language words increases IH interference. It differs from the three Dutch studies in showing that IH interference can arise in the absence of pure non-target language words. It goes beyond the three Dutch studies by showing that adaptive control is possible. IH interference in the Informed+German condition decreased in phase 2 towards the level in the Informed-German condition. In fact, there was a marked difference between these two conditions in the degree of IH interference in phase 1 [$F_1(1,30) = 12.48$, MSE = 26115.46, $p = 0.001$; $F_2(1,76) = 10.17$, MSE = 15667.83, $p < 0.01$; min $F'(1,93) = 5.60$, $p < 0.025$], but no statistical difference between the two conditions in the degree of IH interference in phase 2 [$F_1(1,30) = 1.23$, n.s.; F_2 and min $F' < 1$]. Experiment 4a also provided evidence that individuals adapt not by decreasing the activation of the German system (internal account), nor by reconstruing their task (de Groot, Delmaar & Lupker 2000), but by modifying how they respond to the signals coming from the bilingual lexico-semantic system (external account). Experiment 4b provides a further exploration of the external control mechanism.

Experiment 4b:
Implications of task instructions for controlling IH interference

Individuals informed of the presence of IHs may adjust how they respond to words by, for example, lowering their criterion for responding ‘yes’. If this is the case, then bilinguals informed of the presence of IHs should show less IH interference compared to bilinguals not so informed. This prediction was examined by contrasting the performance of two bilingual conditions, neither of which were shown pure German words. Bilinguals in the *Informed* condition were informed about the presence of IHs in the stimulus list before starting phase 1. Bilinguals in the *Delayed-informed* condition were informed about the presence of such stimuli only before starting phase 2.

This design allows (at least in phase 1) to assess the effects of being informed (which, it was discussed in the Introduction to this chapter, could have been a confound in the study of Dijkstra, van Jaarsveld & ten Brinke, 1998, Experiment 2, and de Groot, Delmaar & Lupker, 2000, Experiment 3). Secondly, varying *when* individuals are informed about the presence of IHs, makes it possible additionally to assess the extent to which individuals can intentionally reduce interference once they have gained experience in a task. If they can control interference in such circumstances, then any difference between the Informed and Delayed-informed conditions in the size of the interference effect which may be present in phase 1 (assuming that being informed affects processing), should disappear in phase 2.

In terms of the locus of such control, if being informed allows to reduce IH interference, and IH interference is reduced by decreasing the activation of the German system (or the units coding for language membership), then the strength of signals indicating the presence of a German word will be less (internal account). In consequence, there will be less activation of the units underlying the 'no' response, and so there will also be less residual activation in the 'no' response units. As a result, the internal account predicts that carry-over effects will be less in the Informed condition than in the Delayed-informed condition, at least in phase 1. In contrast, if control is achieved at an *external* locus, there is no reason to suppose any change in the activation of the German system, and so carry-over effects will be comparable for the two conditions in both experimental phases, even though IH interference is expected to be less in the Informed condition compared to the Delayed-informed condition (at least in phase 1). As one may see, carry-over effects once again play a central role in differentiating between the internal and external accounts.

Experiment 4a also left open the empirical consequences of certain minor manipulations. A subsidiary aspect of Experiment 4b was to examine the sensitivity of the interference effect to these manipulations.

First, would merely mentioning an exemplar IH in the instructions be sufficient to increase the activation of the German system and so increase interference? According to Grosjean (e.g., 2001), this should be possible. To examine this question, the performance of two groups of Informed participants was contrasted. The Informed-German condition of Experiment 4a (who had been shown an exemplar IH in their instructions) was used as one group (called 'Informed$_{+Ex}$'[72] - N = 16). A new group of Informed bilinguals (N = 16) performed the identical task with identical items, but with instructions that gave no IH exemplar (called 'Informed$_{-Ex}$'[73]).

Second, could German nonwords affect the activation level of the German system (in the absence of pure German words)? German nonwords such as PFITZE (orthographically illegal in English) could potentially function in the same way as pure German words appear to do, namely to boost activation of the German system and so increase IH interference. The Informed-German condition's data (Experiment 4a) had suggested that this was not the case (no increase of IH interference from phase 1 to phase 2; German nonwords were shown only

---

[72] 'Informed plus exemplar'

[73] 'Informed minus exemplar'

in phase 2). However, practice might have played a role here, and so this issue was to be addressed further. To assess this possibility, there were two groups in the Delayed-informed condition. One group (N = 16) was shown such German nonwords during phase 1 ('Delayed-informed$_{P1}$'). The other group (N = 16) received them only as of phase 2 ('Delayed-informed$_{P2}$').

5.5 Method

5.5.1 Participants

64 German/English proficient bilinguals (32 male, 32 female), with a mean age of 30.3 years (*SD* = 5.1, range = 20-40), from the same source as participants in Experiment 4a, took part. They were assigned at random to the conditions of the study, subject to ensuring an equal number of males and females in each condition. On average, they had been in the UK for 6.2 (3.6) years. All had normal or corrected vision. As noted above, the data from the Informed-German participants (Experiment 4a) were combined with a new set of 16 bilinguals to form the Informed condition of the present experiment. The 32 participants for the Delayed-informed condition were all new. Participation was voluntary, and was remunerated at the standard departmental rate.

Language proficiency:

Bilingual participants grew up with German and started to learn English at school as their second language at an average age of 9.5 years (2.4). Word recognition scores on the LLEX test (Meara, 1994) averaged 91.2 (4.2) out of 100 (range 79-99) for English and 91.9 (6.1), range 70-99, for German, indicating comparable proficiency in the two languages on this test.

Bilingual participants estimated (language background questionnaire, Appendix 1) that they used English more commonly than German [71.5 (20.0), out of a combined total of 100 for both languages] but considered themselves relatively more proficient in German [54.5 (6.8), out of a combined total of 100 for both languages]. On the scales adapted from Bachman & Palmer (1989; see Appendix 2) their mean scores (out of a maximum of 28) were 25.8 (2.0) and 24.1 (2.3) for German and English, respectively.

5.5.2 Design

Bilingual conditions:

Interference effects. As in Experiment 4a, the data for these analyses come from the IH and control trials. Condition (Informed/Delayed-informed) was a between-subjects factor, Item Type (IH / Control) and experimental Phase (phase 1 / 2) were within-subjects factors. There were 32 participants per condition.

Carry-over effects. As in Experiment 4a, the data for these analyses come from the HF1 and HF2 trials. The design for examining carry-over effects was identical to that of Experiment 4a.

Trial sequences:

Trial sequences were identical to those of Experiment 4a, except for the items in the 'no' positions in the critical sequence. For both groups in the Informed condition, and one of the groups of the Delayed-informed condition (Delayed-informed$_{P2}$ group), both 'no' items were English nonwords (e.g., SWERM) in phase 1. In phase 2 the first of the two 'no' trials was a German nonword (e.g., PFITZE), the second an English nonword. The other group in the Delayed-informed condition (Delayed-informed$_{P1}$) saw German nonwords from the start of phase 1. For this group, the first of the two 'no' item was a German nonword, followed by an English nonword. See Table 5.5

Table 5.5

Experiment 4b. The sequential order of the trials in the critical sequences for each condition/group and experimental phase.
(Note: the Informed$_{+Ex}$ together with the Informed$_{-Ex}$ group make up the Informed condition. The Delayed-informed$_{P1}$ group together with the Delayed-informed$_{P2}$ group make up the Delayed-informed condition.)

	Critical sequence trials in experimental phase 1								
Informed$_{+Ex}$	E_{NW}	E_{NW}	IH	E_{HF1}	E_{HF2}	E_{NW}	E_{NW}	Control	E_{HF1}	E_{HF2}
Informed$_{-Ex}$	E_{NW}	E_{NW}	IH	E_{HF1}	E_{HF2}	E_{NW}	E_{NW}	Control	E_{HF1}	E_{HF2}
Delayed-informed$_{P1}$	G_{NW}	E_{NW}	IH	E_{HF1}	E_{HF2}	G_{NW}	E_{NW}	Control	E_{HF1}	E_{HF2}
Delayed-informed$_{P2}$	E_{NW}	E_{NW}	IH	E_{HF1}	E_{HF2}	E_{NW}	E_{NW}	Control	E_{HF1}	E_{HF2}
	Critical sequence trials in experimental phase 2									
Informed$_{+Ex}$	G_{NW}	E_{NW}	IH	E_{HF1}	E_{HF2}	G_{NW}	E_{NW}	Control	E_{HF1}	E_{HF2}
Informed$_{-Ex}$	G_{NW}	E_{NW}	IH	E_{HF1}	E_{HF2}	G_{NW}	E_{NW}	Control	E_{HF1}	E_{HF2}
Delayed-informed$_{P1}$	G_{NW}	E_{NW}	IH	E_{HF1}	E_{HF2}	G_{NW}	E_{NW}	Control	E_{HF1}	E_{HF2}
Delayed-informed$_{P2}$	G_{NW}	E_{NW}	IH	E_{HF1}	E_{HF2}	G_{NW}	E_{NW}	Control	E_{HF1}	E_{HF2}

Informed$_{+Ex}$ = with exemplar IH in instructions, Informed$_{-Ex}$ = without exemplar IH in instructions
Delayed-informed$_{P1}$ = German nonwords as of phase 1, Delayed-informed$_{P2}$ = German nonwords as of phase 2
G_{NW} = German nonword, E_{NW} = English nonword
IH = Interlingual Homograph, Control = matched control word (matched to English reading of IH)
E_{HF1} = first English HF word following the critical items (IH/control), E_{HF2} = second English HF word following the critical items (IH/control)

5.5.3 Materials

The IH, control and HF words in the critical sequences were identical to those used in Experiment 4a. Again, there were 50% 'yes' and 50% 'no' items. The division of item types for 'no' responses was 44.2% English nonwords and 5.8% German nonwords for both groups in the Informed condition as well as the Delayed-informed$_{P2}$ group. For the Delayed-informed$_{P1}$ group, 'no' responses were 41.3% English nonwords and 8.7% German nonwords.

5.5.4 Apparatus & Procedure

The apparatus was the same as in Experiment 4a. The procedure was identical to Experiment 4a, except that individuals in the Delayed-informed condition were informed about the presence of IHs only before the start of phase 2 (the Informed condition of Experiment 4a had been informed before the start of phase 1).

5.6 Results and Discussion of Experiment 4b

Analyses were again on correct median RTs; error analyses were again on arcsine transformed error proportions. Average RTs reported in the text and tables are mean of median ('M') of the correct subject data. Over all bilingual participants, there were 4.9% errors on IHs and 0.4% errors on controls.

5.6.1 Assessing the influence of the minor manipulations

Preliminary analyses showed that there were no performance differences within the Informed condition between the group who saw an exemplar IH in their instructions ('Informed$_{+Ex}$') and the new group which was not shown an IH exemplar ('Informed$_{-Ex}$'). There was no significant main effect of group, nor any reliable interaction between 'group' and the other variables [all Fs < 1]. Likewise, in the Delayed-informed condition, there was statistically no difference between the group receiving German nonwords from the start of phase 1 (Delayed-informed$_{P1}$) and the group who received them only in phase 2 (Delayed-informed$_{P2}$). Neither the main effect of group, nor the interactions of this factor with the other variables, were reliably significant [all Fs < 1]. Fortunately, the effects of interest are not susceptible to these minor manipulations. Accordingly, Table 5.6 presents the average RT (and errors) to IHs and their matched controls as a function of condition (Informed / Delayed-informed) and phase (1 / 2), but does not differentiate between the groups making up the Informed and Delayed-informed conditions. The analyses on bilingual participants likewise do not differntiate between the two respective groups.

Table 5.6 shows the unilingual baseline group (from Experiment 4a) for comparative purposes. Once again, overall, bilingual participants, despite their high level of English proficiency, were slower, though not more error prone, than the unilingual participants.

Table 5.6

Experiment 4b. Mean of median correct RT (ms), standard deviations (in brackets) and error percent of IHs and controls, as a function of condition/group, phase and item type.
(IH = Interlingual Homograph).

	Phase 1				Phase 2			
	IH		Control		IH		Control	
Condition / group	**RT (*SD*)**	**% Error**	**RT (*SD*)**	**% Error**	**RT (*SD*)**	**% Error**	**RT (*SD*)**	**% Error**
Bilinguals: **Informed**[1]	806 (159)	5.3*	743 (128)	0.7	793 (162)	3.1	718 (141)	0.3
Bilinguals: **Delayed-informed**[1]	838 (183)	5.9*	720 (122)	0.3	839 (238)	5.3*	700 (87)	0.3
Bilinguals: **‡Delayed-informed$_{P2}$**[2]	810 (179)	5.6	720 (137)	0.0	836 (232)	6.9	701 (74)	0.0
Unilinguals[2]	632 (94)	5.0	605 (90)	1.3	622 (82)	4.4	583 (87)	1.3

* Includes 0.6 knowledge errors (i.e., errors on items not known as English words). [1] N = 32; [2] N = 16
‡ Analytical comparison of the Delayed-informed$_{P2}$ group and Unilingual condition follows.

5.6.2 IH interference effects

A mixed-factor ANOVA was performed on the bilingual data. For the subject analyses, Condition (Informed/Delayed-informed) was a between-subjects factor, while Item Type (IH/control) and experimental Phase (phase 1/2) were within-subjects factors. For the item analyses, Item Type was a between-subjects factor.

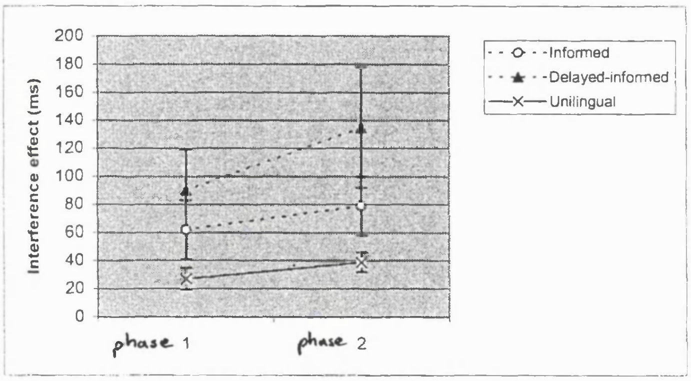
Table 5.6 suggests the presence of an IH interference effect: RTs on IH trials were consistently slower than RTs on control trials. This pattern appears to be more pronounced in the Delayed-informed than the Informed condition, and to not change much from phase 1 to phase 2. (The unilingual data are discussed later on.) Analysis confirmed the visual impression. The principal findings are reported here; further details are given in Appendix 8.

There was a significant effect of item type [$F_1(1,62) = 70.69$, MSE = 8837.39, $p < 0.001$; $F_2(1,76) = 34.83$, MSE = 9082.73, $p < 0.001$; min $F'(1,131) = 23.33$, $p < 0.001$], such that RTs were slower to IHs than to controls. As predicted by the external account, there was also a significant two-way interaction between condition and item type (a trend on min F') [$F_1(1,62) = 6.33$, MSE = 8837.39, $p < 0.025$; $F_2(1,76) = 4.17$, MSE = 7121.56, $p < 0.05$; min $F'(1,136) = 2.51$, $p = 0.11$]. IH interference averaged 129ms in the Delayed-informed condition and 69ms in the Informed condition. There was no reliable three-way interaction between condition, item type and phase [Fs < 1], indicating that individuals informed of the presence of IHs only at the start of phase 2 (the Delayed-informed condition), were unable to reduce IH interference down to the level of the Informed condition. See Figure 5.2.

Note, even the condition (initially) uninformed about the presence of IHs showed an IH interference effect in their uninformed phase (phase1); and yet participants did not seem to have noticed the presence of IHs when asked at the end of the experiment. This is not in accord with the notion that delay on IHs was due to having noticed IHs and hesitating, uncertain how to respond. The pattern is, however, in accord with the notion of two, conflicting, language signals that give rise to response competition.

Figure 5.2

Experiment 4b. IH interference effect (ms), as a function of Condition (Informed / Delayed-informed) and Phase (1 / 2). (Analyses including the Unilingual baseline condition follow).



In terms of errors, there was a significant effect of item type [$F_1(1,62) = 81.18$, MSE = 0.01, $p < 0.001$; $F_2(1,76) = 18.43$, MSE = 0.02, $p < 0.001$; min $F'(1,107) = 15.02$, $p < 0.001$], but there were no more errors in the Delayed-informed condition than in the Informed condition [$F_1(1,62) = 1.76$, n.s.; F_2 and min F' < 1].

Bilingual participants and the unilingual participants (from Experiment 4a) were also compared, using the group from the Delayed-informed condition who saw the *same* stimulus materials as the unilingual participants throughout the entire experiment (Delayed-informed$_{P2}$ - N = 16; see Table 5.5 for the corresponding data and Appendix 8 for details of this analysis). Critically, the interaction between item type and group was significant ($p < 0.01$ by subjects; $p < 0.001$ by items). The RT difference between IHs and their matched controls was greater for the bilinguals than for the unilinguals (see Table 5.6).

5.6.3 Carry-over effects

As in Experiment 4a, RTs to the two high frequency words (HF1 and HF2) following each IH and its matched control were examined, using the RT (by subjects) data; error analyses were again ruled out due to low error rates. A mixed-factor ANOVA was carried out on these data with condition (Informed/Delayed-informed) as the between-subjects factor and prior item type (following IH / following control), position (1/2) and experimental phase (phase 1 / phase 2) as the within-subjects factors. The calculated carry-over effects for Experiment 4b are displayed in Table 5.7. (As in Experiment 4a, for each participant the median RT to HF1 following a control word was subtracted from the median RT to HF1 following the associated IH; the same applies to calculating carry-over on HF2)

As in Experiment 4a, RTs were faster in phase 2 compared to phase 1 [$F_1(1,62) = 19.00$, MSE = 4451.45, $p < 0.001$; phase 1: M = 660ms (113); phase 2: M = 634ms (98)], and there was a significant effect of prior item type. Bilinguals responded more slowly to HF words following an IH (M = 658ms) compared to HF words following a matched control (M = 635ms) [$F_1(1,61) = 34.94$, MSE = 1927.32, $p < 0.001$]. As in Experiment 4a, the carry-over effect decreased significantly from HF1 to HF2 (33ms and 12ms carry-over respectively) [$F_1(1,62) = 10.00$, MSE = 1331.08, $p < 0.01$]. Critically, as predicted by the external, but not by the internal account, the carry-over effect did *not* vary significantly with condition [$F_1 < 1$], while IH interference *had* varied between these two conditions. Such an outcome is incompatible the notion that (informed) individuals exert control by decreasing the activation of the German system (e.g., by inhibition from the language-node level to the word-form level, as the internal account would suggest). If they had done so, then the carry-over effect and the IH interference effect would have decreased together.

Table 5.7

Experiment 4b. Carry-over effect[1] *(based on mean of median correct RT, ms) and standard deviations (in brackets), as a function of condition/group, phase and position.*
(G = pure German word).

Condition / group	Phase 1		Phase 2	
	Position 1	**Position 2**	**Position 1**	**Position 2**
Bilinguals: **Informed**[2]	22 (68)	5 (50)	35 (71)	19 (62)
Bilinguals: **Delayed-informed**[2]	28 (63)	8 (44)	47 (53)	19 (53)
Bilinguals: **‡Delayed-informed$_{P2}$**[3]	48 (68)	15 (44)	54 (66)	8 (67)
Unilinguals[3]	3 (37)	13 (54)	15 (62)	21 (63)

[1] Carry-over effect: calculated as [RT to high frequency words following IHs] minus [RT to HF words following controls], scored separately for the first HF word following IH/control (position 1) and the second HF word (position 2).

[2] N = 32; [3] N = 16

‡ Analytical comparison of the Delayed-informed$_{P2}$ group and the Unilingual condition follows.

Errors were less than 0.8% in each condition, and did not change as a function of phase, previous item type or position; they are for this reason not included in Table 5.7.

Note: Table 5.7 only presents the carry-over effects themselves to facilitate reading the table. Relevant mean RTs for HF words following IHs and those following controls are cited in the text.

As in Experiment 4a, the extent to which bilinguals demonstrated carry-over, whereas unilinguals did not, was examined. To this end, the data of those participants in the Delayed-informed condition who saw the *same* stimulus materials as the unilingual participants throughout the entire experiment (i.e., the Delayed-informed$_{P2}$ group, N = 16; see Table 5.5), were again used.

An ANOVA, with the between-subjects factor group (unilingual / bilingual), and the within-subjects factors prior item type, position and phase, showed that there was a significant third-order interaction between group, prior item type and position [$F_1(1,30) = 7.78$, MSE = 1100.84, $p < 0.01$]. For the HF1, the interaction between group and prior item type was significant [$F_1(1,30) = 7.03$, MSE = 1969.60, $p < 0.025$]. For bilinguals the average RT on HF words following IHs was 671ms, compared with 620ms for HF words following the matched controls. For unilinguals the respective data were 583ms and 574ms (i.e., an average carry-over of 51ms for bilinguals and of just 9ms for unilinguals). In contrast, by HF2, there was no significant interaction between group and prior item type [$F_1 < 1$]. At this point, the effect of prior item type averaged 11ms in the bilinguals (620ms following IHs and 609ms following the controls) and 17ms in the unilinguals (568ms following IHs, 551ms following the controls). These data provide further evidence that bilinguals and unilinguals reacted differently to the IHs.

5.6.4 Responding to German-like nonwords and English-like nonwords

In both the Informed and Delayed-informed conditions, individuals rejected German-like nonwords such as 'UTZT' [M = 786ms (170)] significantly faster than English-like nonwords such as 'SWERM' [M = 842ms (212)] [$F_1(1,62) = 43.84$, MSE = 6304.86, $p < 0.001$; $F_2(1,78) = 35.11$, MSE = 2915.03, $p < 0.001$; min $F'(1,45) = 19.50$, $p < 0.001$], consistent with the notion of using an orthographic strategy for classifying the German-like nonwords as nonwords. In Experiment 4a, in contrast, the 'no' RTs to pure German words (which had no German-specific orthography) had been comparable to those of English-like nonwords. Individuals made no errors on the German-like nonwords and few errors on the English-like nonwords (1.9 % in the Informed condition; 1.6 % in the Delayed-informed condition).

5.7 General Discussion

Experiments 4a and 4b confirmed that German/English bilinguals are significantly slower to respond to IHs than to matched control words in an English-specific lexical decision task. The difference in reaction time for these two types of words exceeds that shown by a unilingual, native-English, control group. This result alone could be regarded as a scaling effect. That is, the subjective frequency difference between IHs and controls (lower subjective frequency for the IHs) could have been exaggerated for the bilinguals, who are after all not native English speakers. However, there are several reasons to believe that the IH/control difference was not just a scaling effect. a) *IH/control differences:* Bilinguals in the Delayed-informed condition demonstrated a larger IH/control difference than bilinguals in the Informed condition (Experiment 4b). There is no plausible reason why a scaling effect should be greater in one randomly assembled group of participants compared to another. b) *Carry-over effects: i)* Carry-over was found only for bilinguals, not for unilinguals. Especially relevant here is the comparison between the unilinguals and the bilinguals who saw exactly the same stimuli (the Delayed-informed$_{P2}$ group; Experiment 4b). *ii)* The carry-over effect was larger when there were pure German words in the stimulus list than when there were none. Furthermore, it remained larger even when IH interference diminished (Informed+German, Experiment 4a, phase 2 vs. 1). *iii)* In the absence of pure German words, carry-over was comparable even when the degree of IH interference differed (Informed vs. Delayed-informed, Experiment 4b). In sum, instead of a scaling effect, the data pattern is consistent with the idea that IH interference arises because of competition between the units underlying a 'yes' response and the units underlying a 'no' response. In an English-specific lexical decision task, interlingual homographs such as MUTTER elicit a German signal that activates the units underlying a 'no' response. Matched control words do not. This experiment was in accord with the assumption that a bilingual's access to his two languages is non-selective.

Experiments 4a and 4b also provided evidence that the degree of interference can be modulated. Interference was greater in the presence of pure German words than in their absence, but only initially (phase 1, Informed+German, Experiment 4a). Such an outcome suggests that pure German words increased the activation of the German system (or the units coding for language membership) and so increased the strength of the signal indicating the presence of a German word. The fact that individuals were able to reduce the degree of interference in phase 2, indicates that a degree of control is possible. One possible explanation for such adaptation derives from the instructions themselves. Experiment 4b confirmed that informing individuals of the presence of IHs before the start of phase 1 decreased the degree of interference. However, such instructions proved ineffective when given *after* phase 1. It is as if individuals were unable to change their mode of responding.

The pattern of IH interference on its own is compatible with two different claims about how bilinguals control their mental lexica in visual word recognition. They might control the activation of the German system (or the units coding for language membership) [internal account], or they might alter how they respond to signals concerning language membership [external account]. Examining carry-over effects made it possible to differentiate between the two accounts. Of principal interest was the difference in RT to an HF English word directly following each IH and its matched control word.

Overall, in line with expectation, bilingual participants showed a carry-over effect whereas unilingual participants did not. Focusing now on the data of the bilingual participants, Experiment 4a showed that the carry-over effect was greater in the presence of pure German words than in their absence (Informed+German vs. Informed-German). Critically, this increased carry-over effect did not diminish together with the decrease of IH interference (Informed+German, phase 1 vs. 2). In Experiment 4b IH interference varied as a function of instructions, but the carry-over effect was invariant (Informed vs. Delayed-informed).

How should the patterns of carry-over effect and IH interference be interpreted? Conceivably, the carry-over effect simply reflects processing difficulty and carries no implications for the nature of control. This possibility can be ruled out because it predicts, contrary to the findings, that the carry-over effect will increase and decrease in line with IH interference and because there should then have been more evidence of carry-over for unilinguals (whose slower RTs to IHs compared to controls presumably reflects some difference in item difficulty).

Consider next the specific pattern of IH interference and carry-over in Experiment 4a. It has been argued here that the increased interference in the presence of pure German words reflects increased activation of the German system or the units coding for language membership (see also Dijkstra, van Jaarsveld & ten Brinke, 1998; Dijkstra, de Bruijn, Schriefers & ten Brinke, 2000), and that adaptation indicates a process of control. Now it might be that the initial degree of IH interference in this condition includes a contribution from an episodic component. Pure German words were presented two trials in advance of each IH and its matched control word. Individuals correctly responded 'no' to such pure German words. That is, in the presence of a +G signal individuals responded negatively. Conceivably then, an episodic trace of this signal-response contingency (cf. Neill, 1997) initially compounded the problem of responding 'yes' to the IH. On this view, adaptation arises because individuals come to dissociate their response to the interlingual homograph from this episodic trace. However, this account then speaks only to the question of IH interference reduction. It cannot explain why the carry-over effect is greater, and remains greater, in the presence of pure German words despite a reduction in IH interference. This divergent data pattern is more readily explained in terms of an increase in activation of the German system induced by the presence of pure German words.

Consider the internal account, in which individuals can inhibit non-target (German) representations. In order to examine this possibility, certain assumptions need to be spelled out. It has been supposed that a correct response is made to an IH when activation in the 'yes' units reaches a criterial level sufficiently above the activation of the 'no' units. On the further assumption that reaching a 'yes' response becomes increasingly difficult as activation in these 'no' units increases, activation in the 'no' units will differentially affect RT to make a 'yes' response on the following trials, even when activation decays from these 'yes' and 'no' units at the same rate. The higher the residual level of activation in the 'no' units, the more difficult it will be for the 'yes' units to boost their residual activation to the criterial level above that of the 'no' units.

The data are in line with these assumptions. RTs to an HF (English) word were slower if it followed an IH than if it followed a matched control (purely English). This effect increased in the presence of pure German words. Such an increase is consistent with greater activation of the units underlying the 'no' response as a result of a stronger signal indicating the presence of a German word. If adaptation in Experiment 4a arose because individuals suppressed the activation of the German system (or the units coding for language membership) [internal account], then the carry-over effect should have decreased as well. It did not. Further, if the reduction in IH interference in Experiment 4b (Informed condition) arose because of reduced activation of the German system, then carry-over effects should also have been reduced. They were not. They were comparable to those of the Delayed-Informed condition.

The more parsimonious explanation is that control is achieved at an *external* locus. An external locus of control has the merit that individuals can modulate IH interference without changing carry-over, as was found. On this account, the size of the carry-over effect reflects the degree of activation of the German system (or the activation of those units coding language membership), induced bottom-up by the presentation of a pure German word. IH interference reflects intentional (top-down) control of how individuals respond to signals from an autonomous word recognition system.

Instructions given at the start of the experimental trials may allow individuals to vary their decision criterion for 'yes' in the presence of conflicting signals (i.e., a German signal indicating a 'no' response, and concurrently an English signal indicating a 'yes' response). Adaptation could arise because individuals shift their decision criterion in the light of the stimuli presented. That is, they could reduce the amount by which activation in the 'yes' response units must exceed that in the 'no' response units. The notion of a flexible decision criterion presumes that individuals are sensitive to the correlation of signals[74] elicited by an interlingual homograph.

5.7.1 The notion of a conjoint test

The notion that individuals are potentially sensitive to the correlation of distinct signals allows an alternative explanation of these data, in which the decision criterion for reaching a 'yes' response (i.e., the criterial amount by which activation in the 'yes' response units must exceed that in the 'no' response units) is comparable for all conditions. On this alternative explanation, as before, a signal indicating a German word activates the 'no' response units

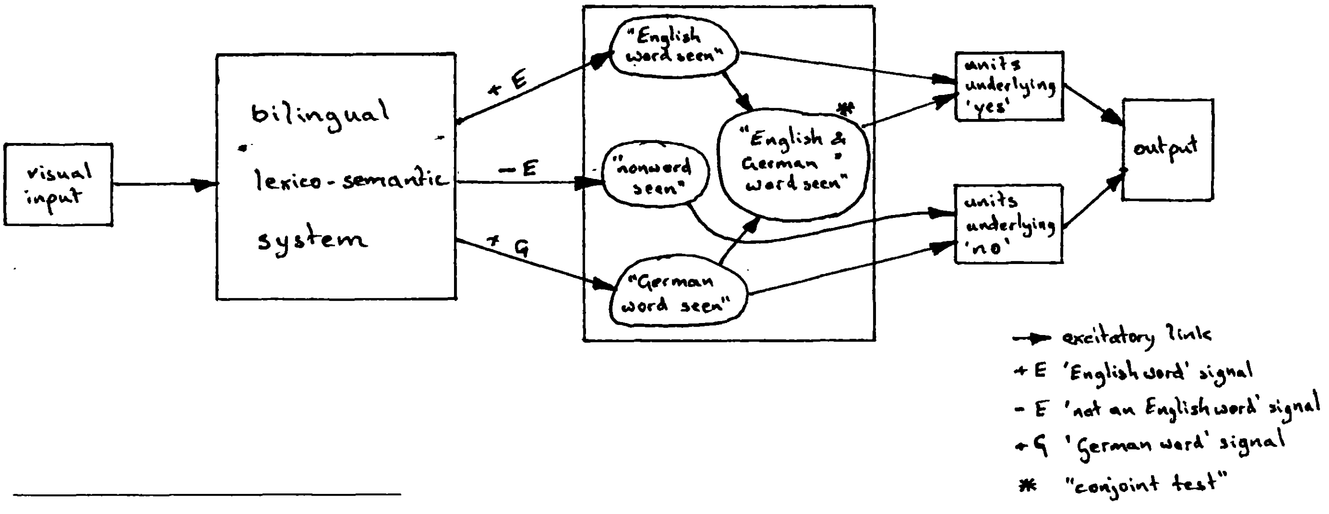
---

[74] a German signal activating the 'no' response units and an English signal activating the 'yes' response units.

and a signal indicating an English word activates the 'yes' response units (because the task schema is an 'English task schema'). However, individuals informed, at the start of phase 1, of the presence of IHs, and how to respond to them, can also establish a "conjoint test". Such a test can activate 'yes', when there is a co-occurrence of a signal indicating a German word and a signal indicating an English word. By feeding activation into the units underlying a 'yes' response, the criterial level of activation of the 'yes' units can be reached earlier. The behavioural consequence is reduced IH interference and a carry-over effect determined by the strength of the German signal, since this German signal continues to determine the activation (and residual activation) of the 'no' response units. Tuning the conjoint test (e.g., by making it increasingly sensitive to English signals) might provide a way in which IH interference could be reduced even in the presence of pure German words[75]. This conjoint test could be envisaged as depicted in Figure 5.3. (A computational implementation is needed to test this conjecture.)

Figure 5.3

Representation of the lexical decision schema showing the conjoint test in an Informed condition. The LD schema for a Delayed-informed condition is identical, but omitting the conjoint test (see text for explanation).



---

[75] If individuals are able to tune their performance, then one would expect them to become increasingly sensitised to signals indicating that a word is English, since this is all-important for determining how to respond in the face of a +G signal. This question was examined in a multiple regression analysis of the IH reaction times in the Informed+German condition (Experiment 4a) using as predictors: Subjective frequency (see *Method* section), phonetic similarity between the English pronunciation and the German pronunciation (cf. Dijkstra, Grainger & van Heuven, 1999) and Coltheart 'N' for English (Coltheart, Davelaar, Jonasson & Besner, 1977), which measures the number of words that can be created from a stimulus word by changing one letter. For phase 1, the overall regression was not significant (F < 1). For phase 2, however, and only for the Informed+German condition, 'N' significantly predicted RT [partial r = - 0.594, $p < 0.01$]. The other factors were not predictive of RT. In sum, in phase 2, individuals in the Informed+German condition, appear to have become more sensitive to 'Englishness' of letter strings.

5.7.2 Connection to other studies and implications

In contrast to Dijkstra, van Jaarsveld & ten Brinke (1998), Dijkstra, de Bruijn, Schriefers & ten Brinke (2000) and de Groot, Delmaar & Lupker (2000), the present experiments obtained, as expected, IH interference in the *absence* of pure German words. De Groot, Delmaar & Lupker (2000) and Dijkstra, van Jaarsveld & ten Brinke (1998, note 4) suggested that, in the absence of pure Dutch words, Dutch/English bilinguals may reconstrue a language-specific lexical decision task as language-general, thus avoiding IH interference. Individuals participating in any of the present experiments do not seem to reconstrue their task even when there are no stimuli (e.g., 'wrong language words') blocking this possibility. 1) IH interference was never eliminated in the absence of pure German words. 2) Reconstrual predicts a pattern of data that was not observed here. It predicts that reductions in IH interference will co-occur with a reduction in carry-over, since a signal indicating a German word will then be linked directly to the 'yes' units rather than to the 'no' units. 3) In the case where pure German words are included (Informed+German condition, Experiment 4a), reduced IH interference (through reconstrual) should be accompanied by an increase in errors on the pure German words (false positives). There was neither reduced carry-over nor an increase of false positives. Individuals participating in the present study (as well as those in the language-specific tasks in Experiments 1a and 2) appear not to have reconstrued their task.

A number of factors may contribute to the difference between Experiments 4a and 4b reported here and the three Dutch studies. Two shall be mentioned here.

First, the participants in this study made, on average, less than 8% errors on IHs, and these were largely processing slips. In contrast, error rates on comparable items were much higher in the three Dutch studies, averaging 19% (Dijkstra, van Jaarsveld & ten Brinke, 1998, Experiment 2), 22% (Dijkstra, de Bruijn, Schriefers & ten Brinke, 2000) and 32% (de Groot, Delmaar & Lupker, 2000, Experiment 3). This could mean that individuals in those studies, relative to individuals participating in the present experiment, emphasised speed over accuracy, but this can only be speculative since these previous studies did not differentiate between processing slips and knowledge errors. Further, the three Dutch studies did not include a unilingual group, so it is impossible to tell whether the bilinguals participating here, despite always being slower than the unilingual group, were in fact relatively slow as bilinguals.

Second, for whatever reason, such as the frequent use of English in a Dutch environment, Dutch/English bilinguals, compared to these German/English bilinguals, may weight the "wordness" or meaningfulness of the stimuli relatively more heavily in reaching a response (cf. Monsell, Doyle & Haggard, 1989; Grainger & Jacobs, 1996). In present terms, a "wordness" signal would activate the units underlying the 'yes' response. Such a signal would be stronger for an IH compared to its matched control word (the former having 'wordness' information from both the English and the German reading). Allowing 'wordness' to bias towards a 'yes' response paralexically would effectively reduce the interfering effects of the non-target language signal that activates the units underlying a 'no' response. On this proposal, the Dutch/English participants in the three Dutch studies performed as instructed, even in the absence of pure Dutch words. It is just that in an English-specific LDT showing no pure Dutch words, they gave more weight to wordness signals than did the German/English bilinguals participating in the present experiments..

These proposals stand in apparent contrast to those of Neumann, McCloskey & Felio (1999) who argued that bilinguals can intentionally reduce the activation of a non-target language (as also proposed by Dijkstra and colleagues). However, their experiment involved a completely different paradigm. English/Spanish bilinguals alternated between two language tasks: naming a word in English and making a lexical decision in Spanish. On the naming trials individuals were required to name an English target word and ignore an English distracter. On the lexical decision trials they were presented with a single letter string for which to make a lexical decision. On some trials they made a lexical decision about a Spanish word that was a translation of the English word named on the immediately preceding trial. On other trials, the Spanish word was a translation of the ignored distracter. On still other trials, it was a Spanish control word that had not been presented for naming. For proficient English/Spanish bilinguals, the switch in language task eliminated the benefit, shown in a unilingual version of the task, of making a lexical decision about words that had previously been named relative to control words that had not been named. It also increased the cost (relative to control words) of reaching a lexical decision about words that had previously been ignored.

Neumann et al. (1999) argued that their data provided evidence for the global suppression of an unwanted language (English) over and above the selective inhibition of distracter words. However, their effects may well be a product of item-specific inhibition and/or of strategic processing. For instance, both languages are active in their task and bilinguals must suppress competing responses not only to avoid naming the English distracter on the naming trial, but also to avoid naming their Spanish translations. Proficient bilinguals are at greater risk from these sources of competition and so may need to impose greater inhibition on non-target items (cf. Green, 1998a; who suggests inhibition of non-target lemmas in production). Further, in order to demonstrate the global suppression of English it is necessary, using catch trials, to show that individuals are slower to perform a task, such as reaching a semantic decision, about English words relative to Spanish words. Nonetheless, their work points to the need to directly address the circumstances that may elicit different loci of control.

The data presented here also speak to the language mode account of Grosjean (1997, 1998, 2001). This connection is a way to explore the implications of this study for more everyday language tasks. Grosjean developed the concept of language mode in order to, for instance, account for the ability of bilinguals to speak in just one language to a unilingual speaker of that language, or to freely code-switch between their languages when conversing with bilinguals who share the same languages. In the former case, Grosjean argues, bilinguals are in a 'monolingual mode', in the latter, in a 'bilingual mode'. These two modes are viewed as a continuum with unilingual and bilingual endpoints, not as two separate states. Variations in language mode are achieved by altering the relative activation of the language systems. Grosjean (e.g., 2001) also applies the idea to reading and, as indicated in the introduction to this chapter, offered an interpretation of the results of Dijkstra, van Jaarsveld & ten Brinke (1998) and Dijkstra, de Bruijn, Schriefers & ten Brinke (2000). Bilinguals can shift towards a bilingual mode as a result of various factors, including the presentation of words from the other language. They can also exert top-down control over their language mode. The locus of control is, in all cases, the activation of the two languages. On the face of it, our results refute this conjecture, since although words from the non-target language increase the activation of that system, adaptation and top-down instructional effects are achieved without altering the relative level of activation of the two languages (IH interference and carry-over effects dissociated). If one takes reductions in IH interference to reflect a shift towards the unilingual end of the 'monolingual-bilingual continuum' (see, for example, Grosjean, 2001), then, according to the data presented here, such a shift is achieved without change in L1 lexical activation.

The results presented here suggest that it is important to distinguish between the lexico-semantic system and the procedures that operate on outputs from this system, or control inputs to it (Dijkstra, de Bruijn, Schriefers & ten Brinke, 2000, also acknowledge this point). Green (1998a, b) referred to these procedures as task schemata. In the present study, instructions required individuals to make use of certain kinds of signals from the lexico-semantic system, in order to make a response. Individuals construct a task schema to perform a lexical decision, mapping language signals and 'wordness' signals from the lexico-semantic system onto responses. Everyday tasks such as reading for meaning, speech production, or translation, require other, more complex, schemata. The general claim that is made here, based on the present data, is that top-down effects (i.e., control) are achieved at an external locus, by the activation of a relevant task schema, rather than at an internal locus, by the activation or inhibition of lexical representations. For instance, in order to read aloud a passage in L1, rather than in L2 (which would effectively be translating it), individuals must activate a reading schema for L1. Such a schema might include directing the eyes to scan from left to right, rather than vice versa; activating L1 rather than L2 phonology; recruiting procedures to build a representation of the text's meaning based on outputs from the lexico-semantic system, rather than procedures to search the text for grammatical or typing errors. Similarly, in order to speak in L1 rather than in L2, individuals must activate the production schema for L1 rather than that for L2. According to Green (1998a, b), these schemata can be configured in various ways. In a unilingual mode, they are mutually inhibited, whereas they may be configured to operate co-operatively if in a bilingual mode, thus permitting code-switching. Mode, it is suggested here, is primarily a matter of which schemata are active, and not so much a matter of which lexical representations are most active. This does not, of course, rule out any influence of the degree of lexical activation: Infrequent use of a language, for example, may lead to a decrease in resting level of activation of the lexical representations of that language.

The data presented here are potentially compatible with Grosjean's proposals, if these proposals are suitably elaborated. For example, the bilingual mode account needs to be expanded to consider what levels of control are needed in order to comply with instructions, or in order to allow for a bilingual's own intentions. On the external account both are expressed by establishing new, or adjusting existing, task schemata.

5.8 Conclusion

The findings presented here are consistent with the notion that German/English bilinguals living in the UK, participating in an English-specific lexical decision task, are in a bilingual mode (e.g., Grosjean, 1998; 2001), and can vary the interfering effects of a non-target language. The findings imply that two distinct loci need to be considered, rather than a single locus based on lexical activation. One locus is within the bilingual lexico-semantic system and is driven by stimulus input. Pure German words (but not, apparently, German nonwords) seem to increase the activation of the non-target system – or, at least, the activation of units coding the items' language membership. The second locus is external to this system and mediates intentional control. Intentional control is achieved, not by altering activation of the non-target system (internal account) , but by changing how signals are mapped onto responses (external account).

The present findings point to the importance of exploring how performance changes during the course of an experiment. These experiments showed that performance adapts, pointing to the ability to control (to some extent) the bilingual language processing system.

The following chapter will provide a review of the experiments presented in this thesis. It will pull together the evidence from each, relating the findings to the hypotheses about the locus of control, critically, of the external and internal accounts.

VI

General Discussion

6.1 Introduction

This thesis investigated how bilinguals control their two languages, an important aspect of fluent bilingual language processing. Some control mechanism must exist, since it *is* possible for bilinguals to translate from one language to another (rather than repeat the words in the same language), or to speak in one given language (rather than the other), or indeed to switch between languages (which has been shown to be not an arbitrary, but a rule-governed process, e.g., Sridhar & Sridhar, 1980). These language acts represent examples from language production, where the necessity to control the languages is most 'visible'. In language perception, the control of languages is less obvious, but just as important. For example, bilinguals are able to understand not only unilingual but also mixed language input. They are able to avoid confusion when confronted with (non-cognate) interlingual homographs (e.g., for German/English bilinguals: "That is very kind" - 'kind' means 'child' in German; or "Look at this mist" - 'mist' means 'dung' in German). For spoken language, phonology generally provides a clear differentiation between languages (e.g., 'kind' is pronounced /kaınd/ in English but /kınd/ in German). For written language, there is no inherent disambiguation on interlingual homographs. Yet bilinguals are able to differentiate seemingly effortlessly. How this may be achieved, that is, what mechanisms are involved in bilingual language processing, is addressed by this thesis.

One phenomenon of bilingual language processing, that has excited much interest, is the language-switch cost. Early studies took this cost as evidence that languages are stored in independent lexica to which there is selective access (e.g., Kolers, 1966; Macnamara & Kushnir, 1971; Scarborough, Gerard & Cortese, 1984; Soares & Grosjean, 1984). That is, the cost for switching between languages was taken as an index of the way in which languages are stored and accessed. It was proposed that a switch had to be set in order to channel stimuli into either one lexicon or the other (see also Penfield & Roberts, 1959). However, this account of the switch-cost is inadequate in the face of evidence (from a variety of paradigms) of interaction between the languages (e.g., Preston & Lambert, 1969 [bilingual Stroop]; Caramazza & Brones, 1980 [bilingual semantic categorisation tasks]; Nas, 1983 [target language nonwords with lexical status in the non-target language]; Guttentag, Haith, Goodman & Hauch, 1984 [bilingual flanker tasks]; de Groot & Nas, 1991 [cross-language repetition- and associative-priming]; Tzelgov, Henik, Sneg & Baruch, 1996 [bilingual Stroop with nonwords that sound like colour names in the non-target language]; van Heuven, Dijkstra & Grainger, 1998 [bilingual word-neighbourhood effects]). In sum, the shortcoming of much of the early research was that experimenters sought to understand phenomena of bilingual language processing (such as language-switch costs and interlingual interference) in terms of the way in which the bilingual lexico-semantic system is structured. The recent approaches, which seek to understand said phenomena in terms of how the system is regulated (be this via language nodes or task schemata), are far more promising as they are able to accommodate previously irreconcilable data patterns.

In this thesis, two old and two new accounts of the language-switch cost were investigated. The two old accounts are the input-switch and the associative connections accounts. The input-switch account (Macnamara & Kushnir, 1971) proposes that incoming stimuli are channelled into one lexicon or the other. Having to re-set this input-switch involves a time cost. The associative connections account (Dalrymple-Alford, 1985) proposes stronger associative connections between the word representations of one language compared to the word representations of different languages. Having to identify a word in language A when language B was previously selected (in the sense used by Green, 1986, 1993), takes time, since the weak cross-language connections do not allow much activation to spread from the language A word representations to the language B word representations.

Beside the two old accounts, two recent ones were evaluated. The importance of these two accounts lies in the fact that they aim to model how languages are controlled. These two accounts differ in their proposition of the locus of control. On the internal account (Bilingual Interactive Activation model, e.g., Dijkstra & van Heuven, 1998), access (to either a common or at least two interdependent representational systems) is non-selective. On this account, processing in the target language leads to greater activation of target language lexical representations (compared to non-target language lexical representations). This relative activation decrement is the means of control: cross-language interference is reduced by inhibiting non-target language representations that could compete with target language representations[76]. A result of this suppression is a temporal cost when switching languages. The degree of inhibition is determined by whether the task at hand requires processing strictly

---

[76] Based on Dijkstra, Grainger & van Heuven's (1999) version of the BIA model that has separate lexical representations for each language (see Dijkstra et al., 1999, Figure 2).

in one language or whether it allows processing in both languages (i.e., language-specific vs. language-general tasks). On this account, the locus of control is within the lexico-semantic system. Deficits of this model are that there is no clear specification of how decisions are made (how processes within the lexico-semantic system are translated into responses), or how control may vary given different task demands - other than by the as yet underspecified suggestion that top-down inhibition (language-node to lexical level) could change according to task demands. How this inhibition could be modulated and what the 'controller' is, has not been expressed.

In contrast, the external account (Inhibitory Control model, e.g., Green, 1998a, b) posits a locus of control outside the lexico-semantic system[77]. This model proposes that during language perception there is *no* inhibition of lexical representations. Note that this account does not specify the lexico-semantic system further. Assumptions about control and response selection are made in terms of task schemata. Task schemata are proposed to react differentially to the signals emerging from the language system during processing. Determined by the task instructions and the targeted level of processing, one or more task schemata are set up and inhibition between schemata is weak or strong. On this model, activation within the lexico-semantic system can be affected bottom-up (e.g., processing in English will raise the activation of English representations), but cannot be controlled top-down (e.g., by inhibiting a given language's representations, as the internal account suggests). On the external account, language-switch costs arise because of competition between task schemata (should there be two schemata, as in the language-specific LDTs) or because of paralexical decision processes, which affect response competition (particularly when there is only one schema, as appears to be the case in the animacy decision task [ADT]).

Neither of these accounts is currently fully specified, making a strong comparison between them difficult. Given the core assumption of the external account (no inhibition of lexical representations), one may wish, instead, to ask: is the inhibitory link from language-nodes to lexical representations, as proposed by the BIA model, *necessary* in order to achieve control? With some - sufficiently specified - control assumption it may not be necessary. In fact, recently Dijkstra & van Heuven (under review, 2002) have distanced themselves from the top-down control view of the BIA model. Their BIA+ model attributes *linguistic*

---

[77] NB, this refers to language perception. For production, the IC model proposes that non-target candidates which emerge during production are inhibited, but not the entire non-target language system; i.e., this inhibition is 'reactive'.

functions to the language nodes (e.g., language coding), but removes non-linguistic, functional, mechanisms from them (e.g., control of cross-language interference). Instead, a *task/decision system*, similar to a task schema level, is introduced to take over the BIA's internal locus of control. The BIA+'s task/decision system goes beyond the IC's task schema: a decision mechanism is specified as part of the task schema (which itself specifies the series of processing steps required for task performance). The decision mechanism is proposed to continuously read out activation levels in the identification system, weighing them to produce a decision (e.g., levels of activation of orthographic / semantic / phonological representations in order to distinguish words from nonwords).

In order to facilitate the description of the two viewpoints put forward by the IC and the BIA models ('there is not / there is an inhibitory link from language nodes to lexical representations'), the terms 'internal account' and 'external account' will still be used from time to time in the following. When they are used, they should be understood with the above proviso. This strategy is pursued in order to facilitate comprehension for the reader who leafs from an experimental chapter to data summary tables and other parts of this chapter, without having read this introduction, which sketches out the limitations of the internal and external accounts and proposes the alternative question whether the BIA's top-down inhibitory link is necessary.

6.2 Review and discussion of findings

There were four main areas of findings: 1) The source of language-switch costs. 2) Cross-language interference effects. 3) The ability of bilinguals to adapt to interference from the other language. 4) The asymmetry of language-switch costs. Since this thesis did not aim to investigate this last point, it shall be dealt with only briefly. The following presents the discussion of how the four different accounts (input-switch, associative connections, internal, external) are able (or not) to accommodate these findings. Summary tables are presented at the beginnings of sections where this is thought to clarify the discussion. Such summary tables list the different accounts and the findings discussed in that section, and signal whether the accounts are able to accommodate those findings.

6.2.1 The source of language-switch costs

The source of language-switch costs was addressed in a variety of ways. Experiments 1a and 1b assessed the question by contrasting the language-switch cost arising in language-specific and language-general lexical decision tasks. The language-specific task preceded the language-general task for all participants who participated in both, and half of the words were high frequency (HF), half were low frequency (LF). Experiment 2 investigated the question further by refining Experiments 1a and 1b. In Experiment 2, language-specific and language-general tasks were also contrasted, but this time stimulus list composition (pure-HF vs. mixed-frequency conditions) and task order (language-specific preceding language-general and vice versa) were additionally manipulated. In Experiment 3, bilinguals were instructed to decide whether the words they were shown referred to an animate or an inanimate entity. All these experiments used the alternating runs paradigm, developed by Rogers & Monsell (1995). That is, the bilingual participants had to switch between languages on alternate trials (e.g., item presentation was English, English, German, German, etc.). The behavioural effect of having to switch languages was not uniform in these tasks. Depending on task instructions and stimulus list composition, the size of the switch cost increased or decreased. The patterns were as follows.

(1) Regardless of stimulus list composition (Experiments 1-2), the cost of switching was significantly greater in the language-specific compared to the language-general LDT. This pattern held true both for the word and nonword data[78].

(2) The cost of switching languages in the language-specific LDT was reduced if this task followed the language-general LDT (rather than when there was no previous language-general task experience), but only if the stimulus list excluded LF words (Experiment 2). Note, this finding was not very robust, but given different predictions about task order patterns for a 'top-down inhibition' account as opposed to a 'no top-down inhibition' account, this finding shall be considered here nonetheless.

(3) When the task required semantic (rather than lexical) decisions, there was a language-switch cost for response-repetition trials, but not for response-switch trials (Experiment 3). Apparently, language information clearly still influenced processing - at least in the first half of the experiment.

(4) In part 2 of the animacy decision task there was no reliable effect of language-switching (for either response-type). In short, it seems that language information does *not* invariably influence decision times.

---

[78] Note, the following shall focus mainly on the word data. Since it is not fully clear what takes place during nonword processing, the nonword data are not easily interpretable. Switch-costs on nonword trials are mentioned separately in section 6.2.2.

How do the different accounts accommodate these findings, if at all?

Table 6.1

Data summary table: The four accounts' ability to predict the language-switch cost patterns.

	(1) bilingual LDT, switch-cost l-g[1] < switch-cost l-s[2]	(2) bilingual LDT, task order effect for l-s[2] task and pure-HF list only	(3) bilingual ADT, no language-switch cost for response-switch trials in part 1	(4) bilingual ADT, no language-switch cost in part 2
Input-Switch Account	x	x	x	x
Associative-Connections Account	x	x	x	x
Internal Account	✓	x	?	x
External Account	✓	✓	✓	✓

[1] language-general
[2] language-specific

6.2.1.1 The Input-Switch account

(1) Greater language-switch costs in language-specific than in language-general LDTs.

The input-switch account has difficulty accommodating this pattern. As discussed in Chapters 2 and 3, it predicts a greater switch-cost in a language-general rather than in a language-specific task. It does so because in the language-specific task there are task cues (colour / position cues) that indicate which language is currently required. This task cue can help to set the input-switch to channel stimuli into the correct lexicon. When there was no task cue (language-general task), there was no other indication of language: Words were orthographically neutral, and the presence of non-unique (orthographically neutral) nonwords obscured the regularity of the language switches. Had participants realised that language switched on alternate trials in the language-general task, they could have used this regularity as a cue for setting the input-switch. As it is, there was, for words, no clear cue to use for re-setting the input-switch on language-switch trials in the language-general task. Hence, in the language-general task, the input-switch would presumably have remained in a given position until there was a clear need to change it. The larger switch-cost in the language-general task (compared to the language-specific task) would have arisen because, on language-switch trials, the letter string would have been searched for in the lexicon last used. Failure to find a representation in the last-used lexicon would lead to the input-switch being re-set and a new search being instantiated in the other lexicon. In contrast, in the language-specific task, the input-switch could be re-set immediately (thanks to the task cue) and then only one lexicon was searched. In sum, given an input-switch account, on language-switch trials, the time taken to determine whether a letter string is a word, would be shorter in the language-specific task compared to the language-general task. Alternatively, if, on language-switch trials, a response was given after the first search was completed but the second lexicon was not searched, there would not be an RT delay, but the response could be wrong.

(2) Reduced switch-costs in a language-specific LDT that follows a language-general LDT (compared to a language-specific task carried out without previous language-general experience), but only for a pure-HF condition.

Apart from predicting a larger switch-cost in a language-general compared to a language-specific LDT, the input-switch account cannot accommodate a switch-cost reduction for only a given task and for only a given condition. If there is a switch-cost reduction in a second task (compared to a first), then the input-switch account would need to explain this as a practice effect. For example, participants could become accustomed to the regime of switching rapidly from one language to the next, and could become faster at re-setting the input-switch[79]. Accordingly, all 'second' tasks should have smaller switch-costs than 'first' tasks. The overall analysis appears to provide evidence in accord with this view (the three-way interaction between trial type, task order and task was n.s.). However, the input-switch account cannot predict that a switch-cost reduction could vary with stimulus list composition. That is, there should have been no difference between any of the conditions with regard to switching and task order. Yet, though not a robust finding, there appeared to be a difference between pure-HF and mixed-frequency conditions with regard to this effect of task order on switch-costs.

(3) In a semantic categorisation task, there are language-switch costs only for response-repetition trials.

Since, on the input-switch account, switching languages requires a re-setting of the input-switch, this account predicts a cost whenever there is a switch of language. However, language-switching in Experiment 3 was significant only for response-repetition trials - and only in the first half of the experiment. This pattern cannot be accommodated by the input-switch account.

(4) In a semantic categorisation task, there are no language-switch costs in part 2 (for either response-type).

When switching languages requires re-setting an input-switch, one must predict a cost

---

[79] While language-switching occurs in normal language use, it does not tend to switch as frequently as every other word, so this certainly might be a factor benefiting from practice.

whenever there is a switch of language. However, language-switching in Experiment 3's part 2 was not significant.

6.2.1.2 The Associative Connections account

(1) Greater language-switch costs in language-specific than in language-general LDTs.

Like the input-switch account, the associative connections account has difficulty with this finding. The assumption underlying the associative connections account is, presumably, that spreading activation underlies switch-costs. With strong connections between word representations of the same language, activation can spread with relative ease among them. The comparatively weaker links between words of different languages means that there is little (or no) spread of activation between languages. In order to 'find' a word's representation (to recognise the word), it must reach a certain level of activation (given a localist model of representation). When a word's representation has already received some activation, it is relatively easier to achieve activation above a given threshold level. On language-switch trials, responses are delayed (compared to language non-switch trials), because there is little or no activation of the word's representation (spreading activation does not affect it much). In contrast, on language-non-switch trials, the target word receives some initial activation via spreading activation. In a language-specific task, the task cues (colour / position) indicate which language will be needed. This can serve to provide some initial activation to all representations of the language cued in this way. In the language-general task, in contrast, there is no task cue, and so some initial activation of the language A word currently being shown is only present if activation from processing in language B on the preceding trial were able to reach it (which, given weak inter-language links, is presumably very little). In sum, on language-switch trials in a language-specific task, the representation of the word shown already has some initial activation, in a language-general task it does not. In this manner, the associative connections account would predict the RT delay on switch trials to be greater in a language-general than in a language-specific task, but this was not shown to be the case in Experiments 1a/1b and 2.

(2) Reduced switch-costs in a language-specific LDT that follows a language-general LDT (compared to a language-specific task carried out without previous language-general experience), but only for a pure-HF condition.

The associative connections account predicted the switch-cost to be greater in the language-general than in the language-specific task (see *(1)* above), and so cannot accommodate a data pattern of this kind.

(3) In a semantic categorisation task, there are language-switch costs only for response- repetition trials.

This finding is problematic for the associative connections account. It predicts a language-switch cost whenever there is a switch of language, because, as discussed earlier, switching languages means having to activate word representations that had little or no benefit from the spreading activation processes that affected word representations of the other language on the previous trial. In contrast to such predictions, the effect of switching languages in Experiment 3 was significant only on response-repetition trials, not on response-switch trials.

(4) In a semantic categorisation task, there are no language-switch costs in part 2 (for either response-type).

The associative connections account predicts a language-switch cost whenever there is a switch of language (see above). However, in part 2 of Experiment 3 there was no reliable effect of language-switching.

Overall, one may see that the input-switch and associative connections accounts are unable to predict either the relative size of switch-costs in language-specific and language-general LDTs (Experiments 1a/1b and 2), task order effects on switch-cost sizes (Experiment 2), or the absence of a language-switch cost on response-switch trials in the ADT (Experiment 3). As shall be described below, the internal and external accounts, which were the main focus of this thesis, are more successful at accommodating these data patterns. The questions addressed shall be both whether a top-down inhibition account can accommodate the data patterns (section 6.2.1.3), but also whether such inhibition is necessary, or whether a different mode of control is plausible, making the notion of top-down inhibition unnecessary (section 6.2.1.4).

6.2.1.3 The internal account

(1) Greater language-switch costs in language-specific than in language-general LDTs.

The degree of top-down inhibition (language-node to non-target language lexical representations) may be presumed to be greater in a language-specific than in a language-general task[80]. The greater the degree of inhibition, the larger the activation decrement that has to be overcome. The consequence is a larger language-switch cost in a language-specific than in a language-general LDT, as indeed was found. In short, the notion of top-down inhibition is in accord with this finding.

(2) Reduced switch-costs in a language-specific LDT that follows a language-general LDT (compared to a language-specific task carried out without previous language-general experience), but only for a pure-HF condition.

Note, the authors of the BIA model have to date not addressed questions of frequency blocking or paralexical decision processes. Hence the following discussion is speculative, based on the existing BIA literature. Note also that the data pattern discussed here was not robust.

A top-down inhibition account has problems with this pattern. Though a 'top-down suppression' account could explain a simple task order effect, it cannot accommodate task order effects varying for different frequency conditions. As a simple effect, a language-switch cost reduction for a 'second' task could be explained with reference to practice. Participants who have learned to choose a setting of weak inhibition of the non-target language during their first task (language-general LDT), could re-call this in their second task (language-specific). In such an event, the switch-cost experienced by a 'language-specific second' group of participants should be smaller than that experienced by a 'language-specific first' group of participants. However, this pattern (smaller switch-cost for the 'language-specific second' group) appeared to apply only to the pure-HF condition, not to the mixed-frequency conditions[81]. The internal account would have predicted all conditions to be comparable, which appears not to be the case.

Can a top-down inhibition account explain the effect in terms of paralexical decision processes? Paralexical decision processes, even though they could be envisaged in the framework of the internal account, cannot explain the switch-cost reduction which individuals in the pure-HF condition appear to have experienced. This is because of the inhibition on

---

[80] Dijkstra & van Heuven (1998): "...the relative activity of the language nodes can in principle be set dependent on the subject's strategy in the task at hand..." (p. 201).

[81] Recall, for the pure-HF condition, in the language-specific task, the switch-cost was 75ms in 'language-specific first' compared to 37ms in 'language-specific second'; but note that this difference, while significant by subjects, was only a trend by items. For the two mixed conditions this trial type x task interaction was n.s. [Fs < 1 or *ps* > 0.2]. Comparing the 'pure' against the 'mixed' conditions (task x task order x trial type x condition), and assessing the 'pure' and 'mixed' conditions separately (task x task order x trial type), suggested that the mixed-frequency condition behaved differently from the pure-HF condition (section *3.3.3* of Chapter 3). However, the finding of the pure-HF condition showing a task order effect while the mixed-frequency condition doesn't, is not very robust.

lexical representations. Specifically, having started at a lower level of activation, it takes longer for a 'general wordness' signal to be generated, thus not providing activation early enough for effective biasing. Accordingly, a top-down inhibition account predicts that there should be *no* difference between the pure-HF condition and the mixed-frequency condition with regard to task order effects. However, Experiment 2 indicated that list composition might be able to modify performance in a language-specific task that follows a language-general one.

(3) In a semantic categorisation task, there are language-switch costs only for response-repetition trials.

It is not entirely certain that this finding can be accommodated given top-down inhibition of lexical representations. Assume that a language-switch invariably leads to a language-switch cost, on the basis of an activation decrement that needs to be overcome when switching languages. Any language signals available early on (e.g., from language-specific orthography) could bias either towards the previously used response (in the case of 'language is unchanged' signals), or towards the response previously not used (in the case of 'language has changed' signals). This would be occurring while an activation decrement is being overcome. Once that is achieved, and the relevant semantic attribute has been ascertained, the corresponding response unit can be selected. If that response unit has been preactivated, it could achieve threshold more quickly. In sum, the 'natural' language-switch cost is decreased on response-switch trials (due to correct bias); by the same token, it would be exaggerated on response-repetition trials (due to incorrect bias). Thus, a pattern as found in Experiment 3's part 1 (see Figure 4.1) could potentially be accommodated by an internal account.

For the account given above, early language signals can create the required biasing effect that exaggerates or decreases the language-switch cost. Late language signals require activation of lexical representations, are thus hampered by the top-down inhibition, and therefore would not have been able to create the observed pattern. For this reason it is not clear whether a top-down inhibition account can accommodate the present findings: the majority of the words used in the present Experiment 3 did not have language-specific orthography, and would thus have provided late, not early, language signals.

To ascertain the ability of the internal account to accommodate the findings, one could contrast two conditions in a bilingual ADT; one whose stimulus list comprised words with language-specific orthography, and one whose stimulus list comprised language-general orthography. If the pattern found in part 1 of the present Experiment 3 occurred only in the 'language-specific orthography' condition, then one could conclude that a top-down inhibition account cannot accommodate the data pattern of Experiment 3's part 1.

(4) In a semantic categorisation task, there are no language-switch costs in part 2 (for either response-type).

A top-down inhibition account has problems with this finding. Given activation decrements of lexical representations and response biases due to language signals, then while a language-switch cost could possibly disappear for one response type (response-switch, see above), there should still be a remaining cost for response-repetition trials. With a certain amendment the internal account would be able to predict the lack of a language-switch cost as was found here: if top-down inhibition (language-node level to lexical level) could be wilfully 'turned off'. In that case, there would be no activation decrement to be overcome on language-switch trials. The authors of the BIA model concede that, theoretically, top-down inhibition could be reduced to zero (T. Dijkstra, personal communication, April 2002). However, if the BIA model needs to allow for top-down inhibition to potentially be 'turned off' in order to accommodate certain experimental results, then that seems to acknowledge that an internal account is insufficient; it requires some external component in order to accommodate the various results known so far.

6.2.1.4 The external account

(1) Greater language-switch costs in language-specific than in language-general LDTs.

This finding can be accommodated without the notion of top-down inhibition of lexical representations. Given language-specific task instructions, a bilingual individual could set up two task schemata with strong mutual inhibition. One links 'English word' signals to 'yes' response units, the other links 'German word' signals to 'yes' response units. Schemata with strong mutual inhibition are required here, as it would be wrong to respond 'yes' to a German word should it appear on an English trial. The currently selected schema strongly inhibits the currently non-selected one. On a language-switch trial, the previously inhibited schema needs to overcome this inhibition, and in turn inhibit the schema that was previously active. This takes time, evidenced as a temporal switch-cost[82]. In a language-general task, there is no need

---

[82] If on a switch trial an error occurs, then this could be due to the task schema 'exchange' processes having been incomplete or ineffective (e.g., Monsell, Yeung & Azuma, 2000). Alternatively, the word could be taken for a nonword (e.g., because it is unknown), in which case the correct task schema might have been established, but the misidentification leads to the error.

for strong mutual inhibition. The time taken to overcome weak inhibition (or to become relatively more active, in the case of jointly active schemata) would give rise to a comparatively smaller switch-cost[83]. Paralexical decision processes may be taking place at the same time as the word identification processes. 'Wordness' signals feeding into the decision process paralexically could bias towards a 'yes' response, which is beneficial to fast responding. Assuming that wordness signals feed into the LD process paralexically, one might assume that the language-switch cost in a language-general task would become so small as to be statistically insignificant. However, 'language-change' signals can presumably also feed into the decision process, and, as discussed in Chapter 4, these signals appear to trigger a change of output. That is, on language-switch/response-repetition trials, 'language-change' signals could delay the (correct) response by biasing towards the incorrect response. Given this, one would expect to find a clear delay on language-switch trials even in a language-general LDT. Whether this suggestion about the role of 'language-change' signals is correct, and whether it is possible to avoid being delayed by such language-change signals, is discussed below (point *(3)*).

(2) Reduced switch-costs in a language-specific LDT that follows a language-general LDT (compared to a language-specific task carried out without previous language-general experience), but only for a pure-HF condition.

A 'no top-down inhibition' account accommodates this finding with recourse to paralexical decision processes. It is proposed that participants confronted with a stimulus list in which all words are HF (pure-HF condition), learn to place much weight on wordness signals. This is in line with the frequency blocking effects found by various researchers in unilingual studies (e.g., Glanzer & Ehrenreich, 1979; Gordon, 1983; Dorfman & Glanzer, 1988). That is, individuals in a pure-HF condition will learn to place much weight on wordness signals if task instructions permit it, i.e., in a language-general task. In a language-specific task, the instructions demand attending to language signals, and therefore, despite a pure-HF list composition, individuals will presumably not adopt a mode in which much weight is attached to wordness signals. It is assumed that previous task experience affects processing in a later

---

[83] The possibility of there being a single task schema in a language-general task, with 'language-change' signals biasing towards the wrong response units and hence inducing a language-switch cost, shall not be ruled out at this stage. However, as discussed in Chapters 1, 2 and 3, it is considered more likely that, in a task that targets the language-coded lexical representations, participants establish separate task schemata, rather than a single schema, even when task instructions permit doing so.

task[84]. In that case, participants are expected to re-call something of what they have experienced before. That is, if they successfully placed much weight on wordness information in a language-general task, then they may do so again to some extent in the following task (language-specific). That is, for a pure-HF condition, wordness signals should play a greater role, and language signals a smaller one, in the language-specific task if it follows a language-general one, than if the language-specific task comes first. With wordness signals playing a greater role than language signals, then on language-switch/response-repetition trials, the correct response units ('yes') can receive early activation (via the paralexical decision processes), and the incorrect response units ('no') receive comparatively little activation. Thus, for a pure-HF condition, in a language-specific task that follows a language-general one, the switch-cost can be reduced compared to when the language-specific task comes first. Note that Experiment 2 concentrated on word trials for this conjecture. Nonwords can be affected by such paralexical signals as well, but this issue shall not be discussed here. Nonwords are discussed separately, with a view to the question of nonword type effects (Experiments 1a/1b and Experiment 2).

In the mixed-frequency conditions no such advantage is expected given a 'no top-down inhibition' account. This prediction was based on the effects of list composition and the strength of language signals. Specifically, it was presumed that individuals in mixed-frequency conditions would be unable to place as much weight on wordness signals as individuals in the pure-HF condition, since they had to avoid rejecting LF words or accepting word-like nonwords. Secondly, if, as Grainger & Dijkstra (1992) suggest, language signals become available later than wordness signals, then individuals who engage in more thorough processing are likely to experience stronger language signals (e.g., a language-node's activation level has a chance to rise more), compared to individuals engaging in less thorough processing. Furthermore, individuals in a 'difficult' condition (mixed-frequency) are likely to try to make the most of all signals available to them, to help them correctly differentiate between LF words and word-like nonwords. Wordness signals will not be very helpful here, since they may be available in comparable measure for LF words and word-like nonwords. In sum, given stronger language signals (and hence presumably stronger 'language-change' signals) and less weight (paralexically) on wordness signals, one would not expect a reduced switch-cost for 'language-specific second' compared to 'language-specific first' in a mixed-

---

[84] It is possible, however, that if the temporal distance between the two tasks exceeds some limit, or there is an intervening event or task between them, that this would not apply.

frequency condition. This is what Experiment 2 indicated.

(3) In a semantic categorisation task, there are language-switch costs only for response-repetition trials.

This pattern can be accommodated without requiring lexical representations to be inhibited. In a task focusing on semantic properties, it is conceivable that bilinguals no longer automatically establish separate task schemata (see Chapter 4). That is, the task instructions do not require separate schemata, and the language-neutral nature of the mental representations (see de Groot, 1998, for a review), increases the chance that only a single schema will be set up. In this case, there should be no cost as far as the task schema level is concerned, but language signals feeding into the process paralexically could still induce a cost. That is, the individual will notice that the language of the input has changed, and such a change of language could be noted as "a change has occurred". This information could be interpreted by the participant as meaning that the current item differs from the preceding one, resulting in a bias (via paralexical decision processes) to the response previously *not* selected. In that case, every time there is a language-switch but no response switch, there is a bias towards the wrong response, causing a delay while the correct response units are activated sufficiently above the ones incorrectly biased towards initially. This correction takes time to achieve, and is evidenced as delayed RT. Indeed, the language-switch cost was significant on response-repetition trials in part 1 but not in part 2, suggesting that initially language signals (specifically: 'language-change' signals) affected responses in this way, but that the impact of these signals was 'tuned down' over time (for example by shifting the weight attached to paralexical signals away from language and more to wordness). (See Figures 4.1 and 4.2.)

To investigate whether the (small) language-switch cost found in the language-general LDTs (Experiments 1b and 2) was due to there being two competing task schemata, as argued here, or whether bilinguals could overcome the force of habit and establish only a single schema, given language-general LDT instructions, may be tested by comparing the size of the language-switch cost in the first half of a language-general LDT with that in the second half. Stimulus presentation would need to be such that language-switching and response-switching were fully crossed (as was done in the ADT [Experiment 3]). If there were a language-switch cost only in the first half, but not in the second half, then this would suggest that it is possible to establish a single task schema even when targeting the language-coded lexical representations, and that bilinguals are able to shift the weight attached to paralexical signals away from language-change signals and towards wordness signals, as Experiment 3 has suggested.

Presumably, if task instructions make it necessary to differentiate between languages in a semantic task (e.g., an ADT that encourages differentiating between languages), then participants would set up two separate and mutually inhibited task schemata. In that case, the external account would predict a constant language-switch cost (regardless of response-type) in an ADT.

(4) In a semantic categorisation task, there are no language-switch costs in part 2 (for either response-type).

This finding is in accord with the external account, which proposed that in a 'language-general' semantic task participants may establish just a single task schema. The pattern furthermore suggests that participants were able to adapt to the signals from their lexico-semantic system, such that 'incorrect biasing' by language-change signals no longer delayed responses. Note, the question of adaptation had not been one of the original aims of Experiment 3, for which reason discussion is deferred to data from Experiment 4 (below).

In sum, the two 'old' accounts have problems with a number of the findings made here. With regard to the two 'new' accounts, the data suggest that top-down inhibition may not be necessary in order to achieve control.

6.2.2 Cross-language influences

The assumption underlying the empirical work here has been that a bilingual's two 'lexica' are not functionally independent, and that access to them is not selective. That is, contrary to researchers like Macnamara & Kushnir (1971), it was assumed that incoming stimuli are not channelled into only one lexicon or the other. Specifically, it was assumed that the perception of a letter string gives rise to activation in both lexica (assuming there are corresponding representations), and that the two lexica interact during processing (cf. for example the cross-language word neighbourhood effects found by van Heuven, Dijkstra & Grainger, 1998).

In this thesis, there were two types of stimuli that could potentially afford evidence of such interaction between the two representational systems ('cross-language influences'). These were, firstly, nonwords, manipulated for their orthographic legality in the two languages (Experiments 1a/1b and 2), and, secondly, interlingual homographs (Experiments 4a/4b). The findings for these two were as follows:

(1) A nonword type effect was found in Experiment 2; in Experiments 1a/1b it was not

statistically reliable.

(2) A significant IH interference effect was found for bilinguals in various versions of the English-specific LDT (Experiments 4a/4b).

Below, these findings are discussed with respect to the question whether or not top-down inhibition of lexical representations is necessary for controlling cross-language interference. Given that the associative connections and input-switch accounts do not predict much (or any) interaction between the two languages, they are not addressed in this section.

The following is divided into two parts. The first deals with nonwords and the question of their showing cross-language effects - and hence whether they can inform the investigation of there being top-down inhibition of lexical representations or not. Previous literature assumed that nonword type effects related to cross-language effects (Altenberg & Cairns, 1983; Thomas & Allport, 2000). It was therefore assumed that patterns of nonword type effects under different task instructions would allow one to distinguish between 'top-down inhibition' and 'no top-down inhibition' accounts (Thomas & Allport, 2000). As discussed below, it is suggested here that this nonword type effect (RT unique < RT non-unique) seems unlikely to reflect cross-language processes, entailing that it does not inform the question whether or not there is inhibition of lexical representations.

The second part is devoted to the assumption the two control accounts are based on - interdependence (rather than independence) of the two lexica. This is discussed in the context of the IH data, given that nonword type effects appear unsuitable for this task.

6.2.2.1 Nonwords

Given evidence of cross-language word-neighbourhood effects[85] (e.g., van Heuven, Dijkstra & Grainger, 1998), one could expect to find nonword type effects in bilinguals (i.e., faster RT on nonwords orthographically legal in one language only ['unique'], than on nonwords orthographically legal in both ['non-unique']); cf. Altenberg & Cairns, 1983. (Thomas & Allport, 2000, included the relevant types of nonwords in their experiment, but did not analyse the nonword type effect as was done here.) Comparatively slower RTs on non-unique

---

[85] 'Neighbours': all words that can be derived from the target word by changing one letter in the same position (Coltheart Davelaar, Jonasson & Besner, 1977; e.g., neighbours of PINE: fine, pane, pile, pins, etc.). Both the numbers of neighbours ('neighbourhood density') and the relative frequency of the neighbours vs. the target ('neighbourhood frequency') influence RT of the target word (e.g., Andrews, 1989; Grainger, O'Regan, Jacobs & Segui, 1989; Grainger & Segui, 1990; van Heuven, Dijkstra & Grainger, 1998).

than unique nonwords *were* found in the relevant experiments of this thesis (Experiments 1a/1b and 2, in both the language-specific and language-general tasks), but this was statistically reliable only in Experiment 2. What could be the reason for this? A simple reason could be that the smaller participant group in Experiments 1a/1b meant that there was insufficient power to establish the effect reliably above the noise of the data. Another reason may be sampling differences (i.e., some participants may have been better in English than others, making it easier for them to confidently reject nonwords). With greater response variation - and a smaller subject group - it may be more difficult to establish a nonword type effect.

Does the presence of a nonword type effect indicate cross-language influences? A nonword type effect could be related to various factors. 1) Non-unique nonwords could activate real-word neighbours in both languages, unique ones presumably only have such neighbours in the target language. When representations in both languages are activated, they could compete with each other, causing delay. 2) Presentation of a nonword string causes activation of that nonword's real-word neighbours. Since the nonword input string only partially matches the neighbouring real-word lexical representations, their activation will presumably only be weak (i.e., not as high as if there were a complete match; e.g., the lexical representation for FINE will become more active if the input is FINE than if it is HINE). When a lexical representation is only partially active (compared to the level of activation it is capable of), it is possible that lateral inhibition is weaker than it would otherwise be. I.e., there would not be much of a cross-language effect at the lexical level, as has hitherto been suggested. The nonword type effect could, however, arise from real-word neighbours having been activated - as follows. As a consequence of activating real-word neighbours, there is general activation in the lexicon. The more real-word neighbours there are, the higher the level of general activation. The stronger this 'wordness' signal, the stronger the activation of the units underlying a 'yes' response. That is, for nonwords with more real-word neighbours, there would be greater bias towards 'yes', and hence greater response competition, than for nonwords with fewer real-word neighbours. Consequently, the former would have slower RTs than the latter. In short, on this explanation, the nonword type effect relates to the response units. 3) As mentioned in Chapters 2 and 3, another possibility is that there is something about the orthography of unique nonwords that allows them to be rejected more easily than non-unique ones. Accordingly, the nonword type effect might even be a very peripheral phenomenon. The data of Altenberg & Cairns (1983) suggest such an explanation since a) they found the nonword type effect both for bilinguals and unilinguals, b) the size of the effect was comparable in the two groups, and c) the bilinguals produced it only in their L1, not in their weaker L2.

What evidence is there in this thesis for the cause of the nonword type effect? There was some indication in Experiment 2 that the nonword type effect found here is more likely to reflect something other than cross-language influences. 1) There was a nonword type x trial type interaction (unique nonwords were 6ms slower than non-unique ones on switch trials, but 32ms faster than non-unique nonwords on non-switch trials). If cross-language influences were the cause of nonword type effects, one would expect them to effect switch and non-switch trials similarly. In contrast, a peripheral orthographic recognition strategy that favours recognition of unique nonwords, could be more effective once a language has been established as *selected*, compared to when its 'selected' status is only just being established. 2) There was an interaction between nonword type and task order. In 'first' tasks, RTs to unique nonwords were only 2ms faster than RTs to non-unique nonwords, while in 'second' tasks this difference increased to 24ms. This is compatible with the notion that bilinguals initially took unique orthography as an indicator that the letter string in question was a word rather than a nonword. This could have been in the form of unique orthography being used paralexically as evidence for a 'yes' response. Then, as participants learned that such orthographic cues were misleading, they could have reduced the weight attached to the orthographic signals, and so, in their 'second' experiment, avoided the problem of having falsely activated 'yes' response units when they were presented with unique nonwords. There is some suggestion that bilinguals used such a strategy mainly in their weaker L2, since error rates on English nonwords were 1.9% greater for unique than non-unique nonwords (ps < 0.01 by subjects and items), while for German there were 1.2% more errors on non-unique than unique nonwords. That is, on German nonwords accuracy favoured the unique nonwords, in contrast to the English nonword trials. Why for German nonwords there should be a higher error rate on non-unique than on unique nonwords is unclear, but in any event, there is no indication that, in the stronger L1, unique orthography was taken as an indication of wordness. 3) Experiment 4a: Individuals in the Informed-German condition saw German-like nonwords (e.g., PFEDE) as of phase 2, but their carry-over effect did not increase compared to phase 1. At the same time, individuals in the Informed+German condition, who saw real German words, produced comparatively greater carry-over. In short, real German words appear to have produced activation in the lexical store, but unique nonwords seem to have been rejected 'peripherally', before lexical representations were activated.

In sum, it is possible that the nonword type effect discussed here reflects some early recognition strategy or response competition, rather than cross-language influences. Given the lack of clarity regarding the source of nonword type effects, it does not seem wise to attach much importance to the main effect of nonword type, and the extent to which it provides evidence for the interdependence assumption. Furthermore, lacking certainty about the origin of the nonword type effect, one cannot use this effect to address the question whether or not there is top-down inhibition of lexical representations. Thomas & Allport (2000) suggested that the internal account would predict a smaller nonword type effect in a language-specific than in a language-general LDT, and rejected the internal account because this prediction did not hold true. This was based on the assumption of cross-language influences, which should be comparatively smaller in the language-specific task, due to the comparatively stronger top-down inhibition. However, as has been discussed above, it is not clear that nonword type effects really do show cross-language influences. Accordingly, in this thesis the lack of an interaction between nonword type and task is not taken as evidence against the notion of top-down inhibition of lexical representations.

To examine the question of cross-language influences (and hence the question of selective / non-selective access and independent/interdependent lexica) one needs to turn to the interlingual homographs. Note, as discussed in chapter 5 (end of section 5.1, section 5.3.3, and section 5.6.2), the delay on IHs (as well as the carry-over effect) is considered to reflect not simply item difficulties, and hence temporarily raised response criteria. Instead, the delay on IHs (as well as carry-over) are considered to be a consequence of the IHs' dual language status and ensuing response competition.

6.2.2.2 Interlingual homographs

Non-cognate interlingual homographs ('IH' for short) are words that have the same spelling in both languages, but different meanings (e.g., MUTTER, 'mother' in German). Given an account of selective access to independent lexica (e.g., Macnamara & Kushnir, 1971), RT and accuracy of responding to such an item should be comparable to that of a control word which exists in the target language only and which is matched to the IH on all relevant parameters.

However, given an account of non-selective access to interdependent lexica, one would expect IHs to behave differently from matched controls. The manner in which they differ depends on various factors. If an IH is presented in a language-general LDT, then the participant could respond 'yes' as soon as he has reliably identified either the one representation or the other. One explanation would be that "...because, given a 'horse-race' model of word recognition, the most frequent reading ... should ... determine the response." (de Groot, Delmaar & Lupker, 2000, footnote 4). Alternatively, evidence from different sources could be pooled (e.g., general activation from both language systems) - the more sources feed into the 'pool', the more quickly evidence accumulates pointing towards the 'yes' response. That is, in a language-general LDT there should be IH *facilitation*. If the same item appears in a language-specific LDT (e.g., English-specific), then this item should give rise to *interference*: The input will activate lexical representations in *both* lexica. While the target language reading gives rise to signals that are linked to the 'yes' response units, the non-target reading gives rise to signals that are linked to the 'no' units. That is, under these circumstances such an item elicits response competition, which must be resolved before a response can be given.

Experiments 4a and 4b used IHs with a higher frequency in the non-target than in the target language reading (increasing the chance of finding interference effects[86], desirable since null effects are more difficult to interpret; see discussion by de Groot, Delmaar & Lupker, 2000). Various manipulations were undertaken, but in each case the task was an English-specific LDT. In all cases, an interference effect was found. However, apart from the IH/control difference that was found for bilinguals (and was called an IH interference effect), there was a small but reliable difference between IHs and controls for the English unilingual group as well. For these people, the slower RT on IHs compared to controls cannot be an IH interference effect, since for them the IH only has one reading (assuming they have not previously encountered these words as German words). This pattern is presumably due to the fact that, as found post-hoc, the English frequency of IHs was lower than that of controls

---

[86] As unilingual ambiguity studies suggest, initially both meanings of an ambiguous word are available. One is then suppressed once it is clear which one is the 'best fit' (e.g., Swinney 1979; Gernsbacher & Faust, 1991; Hino, Lupker & Sears, 1997). This is in accord with the idea of a bilingual IH interference effect: *both* readings are activated, a conflict between them ensues and must be resolved. A bilingual semantic-priming study by Beauvillain & Grainger (1987) indicated that a comparable process takes place with interlingual homographs. At SOAs of 150ms, bilinguals were influenced by the non-target language reading of an IH prime when responding to a target (not an IH). At SOAs of 750ms, this effect disappeared. The present experiments required responses as quickly as possible, rather than asking for responses after a given delay. It is conceivable that interference effects would have been smaller if delayed-responses had been the requirement.

using subjective frequency ratings (using standard ratings [Kučera & Francis, 1967] items had been well matched). Was the IH/control difference in the bilinguals not an IH interference effect either? Was it simply an artefact of frequency, such that IH/control frequency differences were exaggerated for the bilinguals[87]? If that were the case, then there would be no support here for the supposition of cross-language influences and hence of non-selective access. However, as discussed in Chapter 5, there are several reasons to believe that the IH/control difference seen in the bilinguals was not simply a scaling effect:

(1) *Different degrees of IH / control differences:* Bilinguals in the Delayed-Informed condition showed a larger IH/control difference than bilinguals in the Informed condition. There is no plausible reason why a scaling effect should be greater in one randomly grouped set of participants compared to another, who receive identical stimulus lists.

(2) *Carry-over effects*[88]*:* The carry-over effects provide a very clear indication that the IH/control difference experienced by the bilinguals had a different cause than that experienced by the unilinguals. i) Carry-over was found only for bilinguals, but not for unilinguals (this is true for all bilingual/unilingual comparisons, but especially relevant is the comparison between unilinguals and the bilinguals who saw exactly the same stimuli as the unilinguals - the Delayed-Informed$_{P2}$ group in Experiment 4b). ii) Carry-over effects were larger when there were pure German words in the stimulus list (compared to when there were none), and remained unchanged even when IH interference was reduced (Informed+German condition, Experiment 4a, phase 2). iii) In the absence of pure German words (Experiment 4b), carry-over effects were comparable even when the degree of IH interference differed (Informed vs. Delayed-Informed). In sum, not only is the carry-over effect confined to the bilinguals, it is modified by the degree of activation in the non-target lexicon (presumably related to pure non-target words), otherwise it remains unchanged.

---

[87] in the following this is referred to as a 'scaling effect'

[88] Recall, 'carry-over' is calculated as [RT on HF word following IHs] - [RT on HF word following controls].

Table 6.2

Data summary table: The four accounts' ability to predict the IH interference and carry-over patterns.

	(1) different degrees of IH interference for Informed & Delayed-Informed (Exp. 4b)	(2i) carry-over only for bilinguals, not for unilinguals (Exps. 4a & 4b)	(2ii) invariantly larger carry-over effects in the presence of pure German words, despite reduced IH interference (Exp. 4a)	(2iii) similar carry-over despite different IH interference (Exp. 4b)
Input-Switch Account	x	x	x	x
Associative-Connections Account	x	x	x	x
Internal Account	✓	✓	x	x
External Account	✓	✓	✓	✓

These IH/control and carry-over effects are taken as evidence that cross-language influences were found. They are not in accord with the selective access view (e.g., input-switch account, Macnamara & Kushnir, 1971). As far as the question simply of cross-language effects is concerned, it can be accommodated both with and without the notion of top-down inhibition of lexical representations. The dissociation between carry-over effects and IH interference effects is only in accord with the notion that there is *no* such inhibition.

Certainly it would have been preferable if no IH/control difference had been found in the unilingual condition. In future this could be ensured (and hence the issue of scaling effects avoided) by conducting a further pilot study. In a unilingual (English) lexical decision task, IHs and several candidates for matched controls would be presented. All matched controls resembling the IHs in terms of RTs and errors would then be submitted to bilinguals for the rating study undertaken here (to ensure that they would be recognised as English words). The IH and control candidates filtered out in these two procedures would then fulfil the requirement of being well matched for subjective frequency of unilinguals, as well as being known as English words by the bilinguals (recall that IHs and controls were LF in English). Under these conditions, any difference between IHs and controls (in the bilingual data) would then presumably only reflect IH interference, without the uncertainties produced by the potential of a scaling effect. One could, further, counterbalance the presentation of IHs and controls in order to satisfy standard design procedures. As discussed in chapter 5 (end of section 5.2.2), this had not been done in this particular experiment, but was not considered to compromise the data of interest, which was the 'difference of differences' and not just simply the 'difference'.

6.2.3 Bilinguals' ability to adapt to the effects of language signals

The ability to *adapt* to adverse effects of cross-language influences has not previously been addressed in the bilingual literature[89]. Adaptation is visible in the data of two of the experiments presented in this thesis. In Experiment 3 (animacy decision task), this question had not been envisioned at the outset. However, the data indicated that such adaptation can and does take place (language-switch cost on response-repetition trials in part 1 but not in part 2). This question was then directly addressed in Experiment 4a (English-specific LDT showing IHs and including pure non-target language words). The data patterns were as follows.

(1) In the animacy decision task (Experiment 3), RTs were delayed on language-switch/response-repetition trials, but only in the first half of the experiment; in the second half this was no longer the case.

(2) In the Interlingual Homograph (IH) experiment, bilinguals shown pure German words (Experiment 4a; Informed+German condition) showed comparatively greater IH interference than individuals not shown pure German words (Informed-German condition). However, this was only the case in the first half of the experiment. In the second half, there was statistically no difference between the two conditions. At the same time, the Informed+German condition's carry-over effect was greater than the Informed-German condition's, and it remained invariant (while the IH effect decreased).

Does one need top-down inhibition of lexical representations in order to describe this pattern, or can one do without? Assessing this question, what were termed the 'internal' and 'external' accounts so far in this thesis, shall be considered below. The input-switch and associative connections accounts do not predict adaptive control, because they do not presume cross-language influences. They are therefore not mentioned in this section.

---

[89] One study, which at least addresses the question of bilinguals' ability to *control* cross-language interference, is a bilingual colour-word Stroop task undertaken by Tzelgov, Henik & Leiser (1990). They contrasted the case where the majority (80%) of distracter words were in L1, compared to them being in L2. Bilingual participants were imbalanced. Tzelgov et al. found that in the 'high expectancy' (80%) condition, the interference of word-reading on colour-naming was reduced, compared to the 'low expectancy' (20%) condition - but only in L1. This finding was taken to suggest that proficiency involves some ability to control, which is not yet present when the language is 'weak'.

Table 6.3

Data summary table: The internal and external accounts' ability to predict the adaptation patterns.

	(1) adaptation to interfering language signals in the animacy decision task (Experiment 3)	(2) adaptation to IH interference, despite invariant carry-over effects (Experiment 4a)
Internal Account	x	x
External Account	✓	✓

6.2.3.1 The internal account

(1) Adaptation to interfering language signals in the animacy decision task (ADT).

This data pattern can only be accommodated by an account proposing top-down inhibition of lexical representations if it allows for the possibility that this inhibition can be 'turned off'. If this is permitted, and it is shown that the notion of top-down inhibition of lexical representations is not *necessary* for the data patterns of this thesis to be explained, then one might conclude that an account that assumes this inhibition in some situations but not in others is more complicated than necessary.

(2) Reduction of IH interference for the Informed+German condition in phase 2 compared to phase 1, but invariant carry-over.

A reduction of IH interference itself would pose no problems for the notion of top-down inhibition of lexical representations. However, it is incompatible with the accompanying invariant carry-over effect. The reasoning is as follows. IH interference itself can be reduced by increasing the suppression of non-target language word-representations (by increasing the top-down inhibition of lexical representations). The consequence is that any language signals that might emerge from the non-target representational system will be weakened. If they are, then they are less able to provide activation to the units underlying the 'no' response. This will reduce IH interference. However, there will then also be less *residual* activation in the 'no' units on the following trial, and consequently carry-over effects should diminish in line with decreasing IH interference. This was not the case. If, alternatively, bilingual participants reconstrued the language-specific LDT as a language-general LDT (de Groot, Delmaar & Lupker, 2000), then IH interference would be reduced and carry-over would remain invariant, because the activation within the lexicon is then unaltered. However, this is not a plausible explanation for this finding either, because it entails that pure German words (which were shown in this condition), would then be accepted, rather than rejected. That is, error rates on pure German words (false positives) would increase in the second half of the experiment (phase 2) compared to the first (phase 1). This was not the case. Error rates on pure German words remained invariant over the two phases (phase 1: 4.1%, phase 2: 3.8%; n.s. difference). In short, the notion of top-down inhibition of lexical representations as a means of control does not fit with the pattern of concurrent IH interference reduction and invariant carry-over effects.

6.2.3.2 The external account

(1) Adaptation to interfering language signals in the ADT.

Is it possible to account for this pattern *without* inhibition of lexical representations? It seems so. When there is likely to be only one task schema (as in a language-general ADT), a language-switch cost be related to paralexical decision processes. That is, a change of language signal triggers a change of output. Thus, on a language-switch/response-repetition trial, the response which was previously not used, erroneously receives activation. If the individual participating in the experiment succeeds in reducing the weight attached to the language-change signals feeding into the paralexical decision processes, then he can reduce the adverse effects of the language-change signals. Part 2 of the ADT suggested that this is possible, since there ceased to be an interaction between language-switching and response type. In part 2 (where items were translations of items presented in part 1) participants could possibly have recalled an episodic trace of their previous response. This could have helped to reduce the cueing effect of the language-switch. However, if this were the reason for part 2 having no language-switching x response type interaction (as part 1 had), then presumably there should have been fewer errors in part 2 compared to part 1. This was not the case.

(2) Reduction of IH interference for the Informed+German condition in phase 2 compared to phase 1, but invariant carry-over.

The proposal put forward in Chapter 5 was that participants informed about the presence of IHs and how to respond to them would establish not just the simple English-specific task schema which links 'English word' signals to 'yes' units and 'German word' signals to 'no' units, but to incorporate an additional component, a 'conjoint test', that collects English and German signals (see Figure 5.3). When this conjoint test registers co-occurring English and German signals, it activates the 'yes' units. If it registers signals of only one language, it remains mute. Thus, an IH, giving rise to both English and German signals, activates 'no' in that German signals (from the German reading of the IH) are linked to 'no' units, and it activates 'yes' in that English signals (from the English reading of the IH) are linked to the 'yes' units. Given a conjoint test, it can activate 'yes' additionally via the conjoint test route. Accordingly, the more finely tuned the conjoint test is to English signals, the faster or stronger its signals to the 'yes' units can be. This will then reduce IH interference, since additional activation to the 'yes' units speeds up the process of providing 'yes' with sufficient additional activation over and above that in the 'no' units[90]. At the same time, carry-over will remain invariant because the strength of the German signals activating 'no' is unchanged. That is, the residual activation in 'no' units, which hampers the following 'yes' response (for the HF1), is invariant.

In sum, the patterns of adaptation to interfering language signals that were found by Experiments 3 and 4a speak against the notion of lexical representations being inhibited. As regards the lack of a language-switch cost (Experiment 3, part 2) and the combined patterns of IH interference and carry-over effects (Experiment 4a), an explanation for the data found here can be provided by an account that uses task schemata as the means of controlling language signals and hence output.

6.2.4 Asymmetry of language-switch costs

This thesis did not aim to investigate the symmetry (or asymmetry) of language-switch costs. However, since apparently conflicting results emerged in two of the experiments, as well as patterns apparently at variance with data from other studies, the issue needs to be addressed.

Asymmetric switch costs mean that a switch-cost is (significantly) greater when switching from A to B, than when switching from B to A. Such findings have been reported in studies of task-switching (e.g., Allport, Styles & Hsieh, 1994; Monsell, Yeung & Azuma, 2000), as well as in studies of language-switching (e.g., Meuter, 1994; Meuter & Allport, 1999; Thomas & Allport, 2000). Some of these findings have provided evidence that the switch cost is greater when switching from the weaker task/language to the stronger one [weaker stimulus-response mapping → stronger stimulus-response mapping] ('paradoxical' asymmetry; e.g., Allport, Styles & Hsieh, 1994[91]; Thomas & Allport, 2000[92]; Meuter &

---

[90] As discussed in Chapter 5, there is support for this conjecture of increased sensitivity. In a multiple regression analysis Coltheart et al.'s (1977) 'N' (neighbourhood density) was predictive of IH RT - but only in phase 2 and only for the Informed+German condition (the larger the neighbourhood density, the less the delay on IHs).

[91] Experiment 5: Stroop, switching from colour-naming to word-naming

[92] Experiment 2: switching from making language-specific lexical decisions in L2 to making them in L1 (personal communication)

Allport, 1999[93]; Yeung & Monsell, 2000 in preparation[94]). There is also evidence for the opposite, namely that the cost is greater when switching from the stronger task to the weaker one (e.g., Azuma & Monsell, 2000, in preparation[95]; Monsell, Eimer, le Pelley, Stafford & Yeung, 2000, in preparation[96]; Yeung & Monsell, 2000, in preparation[97]). Thirdly, there is evidence that the switch-cost can be symmetrical (e.g., Meuter, 1994[98]; Monsell, Williams, Wright & Rogers, 1996, in preparation[99]; Thomas & Allport, 2000[100]).

In this thesis, the conflicting patterns of switch-cost symmetry were:
(1) Symmetric switch-cost for switching between L1 and L2 (Experiments 1a/1b, LDT).
(2) Asymmetric switch-costs, with greater cost for switching into L2 than into L1 (Experiment 2, LDT).

Two questions arise: 1) How can these findings be reconciled with each other? That is, how can one explain that in one of the LDTs there were symmetric switch-costs while in another they were asymmetric? 2) How can the findings (particularly of (2)) be reconciled with data in the literature which shows that if there is an asymmetry, bilinguals experience more difficulty to switch into *L1* (Meuter & Allport, 1999; Thomas & Allport, 2000, Experiment 2 [Thomas, personal communication]), rather than into L2 (this thesis: Experiment 2)?

In the following, these two questions are discussed with respect to the question whether control of the language system is achieved through top-down inhibition of lexical representations, or whether such inhibition is unnecessary. As shall be shown, the different language-switch cost patterns discussed here can be accounted for without inhibition of lexical representations. The patterns do not speak against this notion, but the inhibition is not necessary; the pattern can be explained without it. The input-switch and associative connections accounts will receive only brief mention, as they are able to accommodate only some of the findings.

---

[93] switching from naming digits in L2 to naming them in L1
[94] switching from tens-complement naming to digit-naming
[95] switching between more and less response-mode-compatible tasks
[96] switching from digit-naming to odd/even classification - key responses
[97] switching from digit-naming to odd/even classification - vocal responses
[98] Experiment 3b: symmetrical switch costs for balanced bilinguals
[99] Stroop, switching between colour-naming and word-naming, using visually degraded words
[100] Experiment 1: switching from making language-general lexical decisions in L2 to making them in L1

6.2.4.1 The internal account

Symmetric and asymmetric (both 'normal' and 'paradoxical') language-switch costs *could* be explained with the notion of top-down inhibition of lexical representations[101]: The L1 language-node is activated bottom-up upon activation of L1 lexical representations. That is, the language-node must be sensitive to activity at the lexical level. If both languages are equally strong (equally practised), both language-nodes are likely to be equally sensitive to the signals from their respective lexical levels. In this case, switch-costs should be symmetric. If, in contrast, L1 ability is dominant, the L1 language-node is presumably more sensitive to activity of 'its' lexical representations. In that case, one would expect to see that it is easier to switch from L2 to L1 than vice versa (i.e., asymmetry favouring L1). A 'paradoxical' asymmetry could also be accommodated. Dijkstra & van Heuven (1998, p. 199) talk of 'asymmetric inhibition', meaning that the top-down inhibition from the two language-nodes may differ in strength. For example, an imbalanced bilingual, to avoid erroneously activating L1 representations and inhibiting L2 representations in a language-specific task, could set a stronger inhibition of the L1 lexical representations (from the L2 language-node) than of the L2 lexical representations (from the L1 language-node). To switch from having a dominant L2 language-node to having a dominant L1 language-node now requires overcoming this additional inhibition on the L1 lexical representations, while switching in the opposite direction would not have this handicap. In that case, the cost for switching would be greater when switching to L1 and smaller for switching to L2. This theory shall now be related to the data.

The data from Experiments 1a and 1b is consistent with this hypothesis: Balanced bilinguals took the same amount of time to switch from L1 to L2 as vice versa. If bilinguals taking part in Experiment 2 were indeed relatively weaker in their L2 than in their L1, then they may have found it easier to switch to L1 rather than to L2, due to different speeds of activating either language-node, resulting in an asymmetry favouring L1, as was found.

Presenting an L1 word on an L2-designated trial would mean that, as part of the recognition process, the L1 language-node becomes dominant. The BIA model does not specify how a lexical decision is taken (how mapping to the output is achieved). One possibility is that having evidence that a language-node has become dominant allows a 'yes'

---

[101] Note, the extant literature related to the internal account does not include any language-switching studies, and has not addressed the question of symmetric and asymmetric language-switch costs. However, it is possible to speculate how it might accommodate this phenomenon.

response to be given. In this circumstance, it would be important to avoid the L1 language node becoming dominant should an L1 word be presented on an L2 trial. This could be achieved by making it more difficult for the L1 language-node to be activated by L1 word-form representations. This then would give rise to a 'paradoxical' asymmetry of switch-costs.

In sum, it is possible to predict symmetric switch-costs (Experiments 1a/1b), 'normal' asymmetric costs (Experiment 2), as well as 'paradoxical' asymmetries (Thomas & Allport, 2000, Experiment 2) with recourse to the notion of top-down inhibition of lexical representations; but is it the only means? The following section suggests not.

6.2.4.2 The external account

The literature on task switching suggests that external stimuli are required to bring a 'new' task schema to dominate (e.g., Rogers & Monsell, 1995). For the present case this means that, for example, an incoming L1 stimulus triggers the L1 task schema, which then inhibits the previously active L2 task schema. If both languages are equally strong, the task schemata may be equally sensitive to signals from either language, such that switch-costs are symmetric. If one language is weaker than the other, it is possible that the corresponding task schema is less sensitive to 'its' signals, and is hence less easily triggered than the stronger language's task schema. It may also be that the signals of the weaker language are available later (e.g., due to lower baseline activation levels / higher activation thresholds of the mental representations[102]). Accordingly, the signals required to establish a task schema (in a bilingual LDT) would be available later for the weaker than for the stronger language. Additionally, with less practice (in the weaker language), the processing system could be less sensitive to the relevant signals. In short, L1 task schemata may be more easily triggered by L1 signals than L2 schemata are by L2 signals, and an asymmetric cost (favouring L1) would occur.

In Experiments 1a/1b language-switch costs were symmetric. In contrast, in Experiment 2, the switch-cost was asymmetric, favouring L1. (Note, the relevant task for this question is the language-specific LDT, since in that case two clearly inhibiting task schemata are required. Accordingly, the following refers to this task, not to the language-general one.)

---

[102] From the PDP point of view: less practised auto-associative links between the attributes of a weaker language word's representation in the input domain, so that it takes longer to reinstate the activation pattern corresponding to a given word (e.g., Allport, 1985; McClelland & Rumelhart, 1985); alternatively, less weight on connections between the input units and the hidden units for the less known words' representations (e.g., Seidenberg & McClelland, 1989) - that is, less effective sets of hidden units for the less known words than for better known words.

The tasks and paradigms were identical, and the stimuli used in the two experiments were essentially the same as well. A possible reason why the data patterns differed despite these experimental similarities, is that the two subject groups differed in terms of their language abilities, with apparently more balanced bilinguals in Experiment 1a. Meuter (1994) has provided evidence that language proficiency plays a role in determining whether switch-costs are symmetric or asymmetric: When bilinguals were imbalanced (stronger in their L1), switch-costs were asymmetric - switching to L1 gave rise to a larger cost than did switching to L2 (Meuter, 1994, Experiment 3a). When bilinguals were balanced, switching between L1 and L2 was symmetrical (Meuter, 1994, Experiment 3b). Qualifying this finding, Thomas (1997, Experiment 1) has provided evidence that relative (not absolute) language abilities determine switch-cost size. In this thesis, on the LLEX test data, bilinguals in the respective experiments appear to have been comparable in their language abilities (German was on average better by 2.3 points in both groups), but on the self-rating data the groups differed. For the participants in Experiment 2 German was considered stronger than English by an average of 15.6 points, while for the participants in Experiment 1a the difference was 14.0 points. This difference is very subtle (and was not statistically significant), but may indicate the reason for the different switch-cost patterns. Had a more sensitive language test been available at the time, it might have been possible to follow up this question more effectively[103].

The second point addressed above was that while in this thesis asymmetries favoured L1, there is evidence in the literature of an asymmetric language-switch cost favouring L2 (Thomas, personal communication, September 2000, referring to Experiment 2 of Thomas & Allport, 2000[104]). Experiment 2 of Thomas & Allport (2000) used the alternating runs paradigm for a bilingual language-specific LDT, as was also the case in the present Experiment 2. However, the two experiments differ in one important respect, which may hold the key to the different asymmetry patterns. Thomas & Allport (2000) included 'wrong language words' (e.g., an L1 word on a trial designated as L2), while the present Experiment 2 did not. To avoid their stronger/more sensitive L1 task schemata being erroneously triggered by 'wrong language' L1 words (which would have entailed the danger of false

---

[103] Meara, the author of the LLEX test, has indicated that for highly proficient bilinguals his '10k' test, which contains a structured sample of words from a lower frequency band (6k-10k), is more sensitive (Meara, personal communication, April 2001). The LLEX, in contrast, comprises a higher frequency range, designed for testing 'Threshold Level' competence, as defined by the Council of Europe.

[104] The question of asymmetry is not addressed in the published version.

positives on such items), Thomas & Allport's (2000) participants may have set a stronger inhibition of the L1 task schemata than of the L2 task schemata. Accordingly, to switch from L2 to L1 would have entailed overcoming stronger inhibition than vice versa, entailing a larger switch-cost when switching into L1, than when switching into L2.

In sum, it seems that these data patterns can be explained without requiring top-down inhibition of lexical representations. Furthermore,, these data suggest that whether symmetric or asymmetric language-switch costs are found seems to be a question of the relative language abilities (this thesis, Experiments 1a / 2; Meuter, 1994, Experiments 3a / 3b). The question of the direction of an asymmetry appears to be determined by stimulus list types (inclusion/exclusion of 'wrong language' words), as suggested by the comparison between Experiment 2 this thesis, and Experiment 2 of Thomas & Allport (2000).

The conjectures regarding the differences between Experiments 1a and 2 (this thesis) - symmetry in the one, asymmetry in the other - could be investigated by comparing a balanced and a clearly imbalanced group of bilinguals, carrying out a language-specific task, using the same stimulus lists. If the language x trial type interaction (i.e., the asymmetry) were modified by the factor group, then this would support the notion that differences in language abilities underlie differences in switching performance.

The speculations regarding the differences between the present Experiment 2 and that of Thomas & Allport (2000), could be substantiated by making a direct comparison between equivalent groups of bilinguals, both of which received a bilingual language-specific LDT using the alternating runs paradigm, but which would differ in that only one group would receive 'wrong language words'.

In either investigation a sensitive language test would be paramount.

6.2.4.3 The Associative Connections and Input-Switch accounts

Following the word association hypothesis of Potter, So, von Eckhardt & Feldman (1984), Kroll and colleagues (Kroll & Sholl, 1992; Kroll, 1993; Kroll & Stewart, 1994) proposed, on the basis of translation studies, that the connections between lexical representations of L1 and L2 differ in strength. That is, they found translations to be faster when translating from L2 to L1 than vice versa (see also Sánchez-Casas, Davis & García-Albea, 1992). This was interpreted as meaning that the connections from L2 to L1 are faster and/or stronger than those from L1 to L2 (which could be due to having learned L2 by analogy to L1).

Adopting such a model, the associative connections account (Dalrymple-Alford, 1985) would be able to predict asymmetric switch-costs when language ability is imbalanced. Depending on the individual's language history, the connections could be of unequal strength, or equally strong (though always weaker than intra-language connections). Relating this idea to the present findings, it could be said that less balanced bilinguals have stronger associative links from L2 to L1 (see Kroll & Stewart, 1994, Figure 3), and hence experience less of a cost for switching to L1 than for switching to L2 (as found in Experiment 2), because stronger links could also allow comparatively more inter-lingual spreading activation than weaker links would. More balanced bilinguals, in contrast, could be regarded as having cross-language links that are comparable in strength, and so the switch costs would be symmetrical (as found in Experiments 1a/1b). However, it is difficult to explain paradoxical asymmetries (Thomas & Allport, 2000, Experiment 2 [Thomas, personal communication]) using the associative connections account. It is not clear how the strength of associative links could be influenced for the purpose of an experiment showing 'wrong language' words.

The input-switch account. On this account, one could argue that re-setting the switch from L2 to L1 is faster than vice versa, perhaps because the recognition processes that identify any available language signals are more finely tuned for L1 than for L2. In contrast, if language abilities are balanced, that means that re-setting the switch takes the same amount of time, whether it is to switch into L1, or to switch into L2. That is, the input-switch account can accommodate the data of the present Experiments' symmetric and asymmetric switch-cost patterns. However, the paradoxical asymmetry found by Thomas & Allport (2000, Experiment 2 [Thomas, personal communication]), poses a problem for this account. To avoid false positives on wrong language words, the imbalanced bilingual might have to avoid feeding such an item into the corresponding lexicon. Could a bilingual inhibit the recognition processes, in order to prevent the input-switch being automatically set towards the 'L1 lexicon' on presentation of an L1 word? It is not clear that it is possible to actively avoid recognising words of a given language.

In sum, in order to account for the asymmetry and symmetry of switching patterns found in this thesis, as well as the paradoxical asymmetry patterns provided by other language-switching studies (Meuter, 1994; Thomas & Allport, 2000), there is no need to inhibit the activation of the 'other' language's lexical representations. The associative connections and input-switch accounts could accommodate symmetric and 'normal' asymmetric switch costs, but have problems with paradoxical asymmetry.

6.3 Conclusion

The principle findings in this thesis that addressed the question of the locus of control, pertained to the origin of language-switch costs and the ability of bilinguals to adapt to interference (reducing its impact). The issue of cross-language effects was addressed as well, since parallel access to interdependent lexica is assumed both by the BIA and the IC model, whose notions of present/absent top-down inhibition of lexical representations was investigated here. Finally, given the evidence in the bilingualism literature of asymmetric language-switch costs, and given that there were some findings regarding symmetry here as well, it was important to assess how these patterns fit with the questioned top-down inhibition (though note that asymmetry was not the focus of this thesis).

The findings made here contribute to the field of bilingual language research in two ways in particular. Language-switch costs were shown to vary with task demands, and for the first time it was shown that adaptation to adverse influences from the non-target language is possible. Is the notion of inhibiting lexical representations, as proposed in the BIA model (e.g., Dijkstra & van Heuven, 1998) *necessary* in order to explain these data patterns, or is it possible to do without it? The discussions in the preceding chapters, and in particular in the present one, suggest that it is possible to account for these data without recourse to such inhibition; the notion of task schemata (e.g., Green 1998a/b) is able to play the role of 'controller'. Indeed, data obtained in the frame of this thesis indicate that while activation levels of the respective languages' lexical representations can be raised exogenously (Experiment 4a: when including pure non-target words, greater IH interference and greater carry-over effects), they cannot be reduced endogenously (Experiments 4a/4b: invariant carry-over despite reduced IH interference). That is, there is, in fact, evidence that speaks *against* the notion of inhibiting lexical representations. Further evidence against this means of control came in the form of the 'missing' language-switch cost in the (language-general) animacy decision task (Experiment 3, part 2).

It seems, then, that although activation within the lexicon is likely to be variable (e.g., more active German than English representations when German is the selected language), this is not the locus of control. The data presented in this thesis suggest that it is important to distinguish between the lexico-semantic system and the procedures operating on the outputs from this system. These procedures have here been characterised as task schemata. This mechanism of task schemata, which operates outside the lexico-semantic system, appears to be influenced by task demands, experience, item constraints and available information. Such a task schema account is parsimonious, proposing the same mechanism for resolving competition between languages, as the mechanism for the control of action (Norman & Shallice, 1986; Shallice, 1994).

This thesis showed that the internal account, as it stands, proposes a locus of control incompatible with various data patterns. One might wish to reject the internal account on such grounds. However, the concept of a language-node should not be discarded outright. It is valuable as a notion of language coding. A language-node, connected to certain lexical representations, and becoming activated over and above another language-node, is a means of coding lexical representations for language and at the same time providing a mechanism for allowing a given language to be 'selected', rather than simply 'active' (using Green's terms [1986, 1993]). What appears to have been refuted here is the notion that such a language-node supplies the locus of control by inhibiting lexical representations of the other language. (See Dijkstra & van Heuven, under review 2002, whose BIA+ model moves in this suggested direction.)

It became clear in the course of this thesis that the interdependence of a bilingual's two language systems can be shown with items that have a representation within the lexicon (e.g., IHs), but not (or at least not easily) with items that do not (e.g., unique vs. non-unique nonwords). That is not to say that nonwords can never achieve this. Work on pseudohomophones (e.g., Nas, 1983; Tzelgov, Henik, Sneg & Baruch, 1996) has provided evidence that a bilingual's two language are not strictly separable, as earlier models (e.g., Penfield & Roberts, 1959; Macnamara & Kushnir, 1971) had suggested.

This thesis joins the existing data in the literature which suggests that a bilingual's two languages are not two separate, independent entities (see also Grosjean 1989), but rather, that the two interact. Furthermore, it provides evidence that bilinguals are able to control these two interacting languages, and goes some way to proposing how such control is achieved.

The discussion of the different findings in relation to the different accounts has highlighted the need to test an account by investigating more than one data sample. For instance, while IH interference and the reduction of IH interference are in accord with the notion of inhibited lexical representations, the dissociation between IH interference effects and carry-over effects (e.g., IH interference reduction despite invariant carry-over, Experiment 4a) is not. The first three experiments of this thesis furthermore highlight the value of testing a hypothesis using different manipulations and different experiments. That is, firstly, while the notion of top-down inhibition of lexical representations as a means of control is compatible with different switch-cost patterns in language-specific and language-general LDTs, it has trouble accounting for task order effects for only one frequency condition but not for another (Experiment 2; though recall that this effect was not statistically robust). Secondly, the internal account was unable to predict the lack of a language-switch cost in a semantic task (Experiment 3, part 2).

The same applies to the input-switch and associative connections accounts. They were able to accommodate some of the data (e.g., the presence of a switch-cost in general, and in particular in a language-general LDT), but not all (e.g., the lack of a language-switch cost in Experiment 3 when switching responses).

So far, the data collected are in accord with the notion that there is *no* inhibition of lexical representations, but that control of the bilingual lexicon could be achieved externally, by task schemata. The task schema account needs to be tested further though, to ascertain whether it can reliably predict other patterns of bilingual behaviour in different situations not investigated here.

What are the implications for 'normal' bilingual language processing of saying that there is no top-down inhibition of lexical representations? For example, if a bilingual wishes to translate German text into English, he avoids reproducing the German text not by suppressing German representations, but by setting up task schemata that allow production to be restricted to English. For spontaneous speech, a bilingual might choose to speak only one language, in which case (as above) a task schema is selected that allows production only in that language. Should the bilingual wish to switch between languages (e.g., when talking to another bilingual), the two task schemata could be in co-operation, rather than one of them dominating. The advantage of this modus of control (i.e., without top-down inhibition of lexical representations) is that for a bilingual situation, both production and comprehension of 'mixed material' can be processed more quickly if there is no need to overcome activation decrements. Indeed, in competent bilinguals, mixed production and comprehension typically proceed in a very smooth and uninterrupted fashion.

6.4 Outlook

Some additional points that need to be considered now pertain to the ecological validity of experimental data, converging evidence from other research methods, and bilinguals' language impairment - given the notion that control is achieved without inhibiting lexical representations of language B when processing in language A.

6.4.1 Ecological validity and converging evidence

Establishing converging evidence is valuable in its own right, but it is particularly important if it means that one obtains data in natural language-usage circumstances, which validates the data obtained under less natural circumstances (such as lexical decision tasks).

The experiments reported here provide data from laboratory settings, collecting data from single word paradigms. During normal language usage, people perceive and produce streams of words, rather than just single words. Furthermore, they use words to convey meaning, not to make lexical decisions about them. In others words, the mental processes which occurred when bilinguals were carrying out the present experimental tasks, may have been different from what normally occurs in language processing. Naturally, the present interest is to shed light on the processes taking place during language processing in general terms. It would not be very sensible to research something which does not happen other than in an experimental setting, unless one were simply interested in the brain's ability to adapt to new circumstances. It is therefore important to look for converging evidence that is able to investigate language processes in as normal a setting as possible. This would augment and validate the RT and accuracy data collected here.

A previously much-debated issue in the bilingual language domain is the question of selective/non-selective access and independence/interdependence of the lexica. It was assumed here that access is non-selective, and that the lexica are interdependent. Evidence for this position came from, for example, the interlingual homograph (IH) experiments (Experiments 4a/4b). Where else might one obtain evidence for this position?

One alternative measure is to record eye movements during reading. The data frequently investigated in eye movement studies are saccade lengths (the eyes' movement from one fixation point to the next), fixation times (the length of time the eyes are focused on a given position before moving away), regressions (re-reading text) and the probability of words being skipped. Eye movement data provides a relatively natural, on-line means of measuring psycholinguistic processes (for a discussion see Rayner & Sereno, 1994). To investigate the question of non-selective access/interdependent lexica, one could present bilingual individuals with unilingual sentences to read, which contain non-cognate IHs (and matched controls). If bilinguals were unable to limit processing entirely to one language (as assumed here), then one would expect to observe different eye movement patterns for IHs than for controls, caused by the IH's ambiguity. On the other hand, if bilinguals *were* able to 'shut off' the non-target language (as Macnamara & Kushnir, 1971, and others suggested), then one would expect to see comparable eye-movement patterns on IHs and controls. That is, using such a technique one would be able to obtain information on the selectivity/interdependence issue using a language task rating highly on ecological validity, as well as obtaining converging evidence.

Another possible method of examining the questions which have been posed here is neuro-imaging. Certainly, RT and error data do not and cannot make any claims as to the location of certain processes within the brain. Converging evidence could and should be sought from studies investigating areas and patterns of activation in the bilingual brain (e.g., by positron emission tomography [PET] or functional magnetic resonance imaging [fMRI]). Some initial work using PET was conducted by Price, Green & von Studnitz (1999). In their study they found that, when translating, cortical areas previously associated with the control of action were involved, suggesting the need for greater mental co-ordination when translating (e.g., pathways associated with reading must be inhibited in favour of the less practised translation pathways). When switching languages, there was increased activation in areas normally associated with orthography-phonology mapping, suggesting that processes related to phonology are a source of difficulty when switching languages. In fact, this corresponds with much anecdotal evidence. Frequently, when bilinguals insert a language A word into an otherwise language B sentence, the word-form and the syntax adjustments (where necessary) tend to be faultless, but the pronunciation of the 'borrowed' language A word is coloured by language B phonology (i.e., the borrowed language A word is spoken with a language B 'accent', even if the bilingual otherwise speaks language A without any accent of language B).

Knowledge about the location of specific processes would be of great interest in order to test the validity of traditional assumptions about language impairment in bilinguals (be this of sudden onset, for example following stroke or head-injury, or of a gradual nature, as in Alzheimer's). On the basis of this knowledge, it would be possible to develop new approaches to treatment where necessary. These issues are considered next.

6.4.2 Implications for bilingual language impairment

Greater knowledge of the anatomical underpinnings of bilingual functioning and malfunctioning may make it possible to assess and treat pathological disorders of language, such as that of the patient AD (Paradis, Goldblum & Abidi, 1982), whose case sparked the interest to undertake this research in the first place. The anatomical structures relating to the lexico-semantic system (the 'data storage' system) and those relating to control of this system may be anatomically distinct (as suggested by the data of Price, Green & von Studnitz, 1999). Accordingly, language impairment following brain injury may be due to the data structures being compromised, or they could be due to dysfunction of the control mechanism (cf. Morton & Frith's 'levels of description' [1995]). For example, it could be that there is an impairment of the allocation of resources, which are required to effectively use the control mechanism. If, for a given aphasic bilingual, it were possible to identify their language impairment as stemming from control problems rather than from data-storage problems (be this for language production or comprehension), it may be possible to alleviate (or perhaps even resolve) their difficulties through drug treatment.

Braver & Cohen (2000) suggest that context representations (which includes goal representations) are kept active by the prefrontal cortex (PFC). Furthermore, "...active context representations serve to mediate control by modulating the flow of information within task-specific task pathways such that processing in the task-relevant pathway is favoured over a (possibly stronger) competing pathway." (Braver & Cohen, 2000, p. 714). In their view, the neurotransmitter dopamine plays a significant role in achieving this. Here, then, is a proposition of a physical substance influencing the ability to control. Individuals suffering from Parkinson's disease experience difficulties with controlling muscle movements (such as hand tremors), related to decreased levels of dopamine. Such individuals also have problems with their speech, which to some extent may be attributed to the deficit of control over the fine muscular movements required for precise articulation. What has apparently not been investigated so far, are language disorders in bilingual Parkinson's patients over and above their muscular deficits. In brief, if dopamine plays a role in the PFC control mechanism, then what effect does a dopamine deficit have on such a bilingual's language processing? Presumably, he would experience the known muscular effects, but additionally experience problems with maintaining control over his two languages. That is, this bilingual's language disorder would not be related to the data storage system, but instead to the control mechanism.

What are the implications for treatment of viewing bilingual aphasia as a result of damage to data storage as opposed to control mechanisms? To date the approach to aphasia therapy has been to teach methods to circumvent the problem (e.g., describing words), re-learn words if possible, and learn to adjust to the new circumstance. For example, should the bilingual aphasic experience problems with finding words in the desired language, then it would be assumed that this is because his output lexicon has been damaged. He would therefore be taught to describe the words he cannot recall, and to accept that he has problems remembering words. Certainly both the therapist and the patient would be frustrated if they experienced the word-finding abilities to fluctuate, without an apparent explanation for this pattern. Besides circumventing techniques, the therapist might attempt to reinstate representations for important day-to-day words (e.g., by presenting words in various modes and contexts). Both the circumventing techniques and the re-learning would be unnecessary, however, if the representations were undamaged, and the deficit were due to problems with control over the language system. Furthermore, if it could be established that the impairments derived from control over an intact data storage system, the pattern of deficits would be understandable and hence less unsettling.

In chapter 1 the alternate antagonism experienced by patient AD (Paradis, Goldblum & Abidi, 1982) was described, and it was discussed that such a pattern of language impairment could not be explained with the traditional view of language impairment (destruction of areas of the brain that underlay the now-missing functions). Taking the notion which has found support from the experiments presented here ("control of the bilingual lexicon does not require inhibition of lexical representations"), one may now attempt to explain the pattern. To be able to produce spoken output in L1, patient AD's "speak in L1" task schema had to inhibit her "speak in L2" task schema. She was able to do this, but was then unable to switch to producing in L2 - until the next day. Such a pattern suggests problems of control (possibly of allocation or availability of resources to affect the control; see Green, 1995).

Once it is possible to distinguish with greater certainty between deficits of data storage and control mechanisms, it may be possible to diagnose that a patient's symptom x is the result of deficits in the ability of a given cognitive process to regulate another cognitive process, and that boosting a certain neuromodulator by administering drug y will alleviate this problem. However, this is not to say that all language processing deficits are curable with the administration of drugs. There are likely to be cases where the anatomical brain systems have been implicated in the damage so as to not be amenable to drug treatment. The aim is to increase the knowledge in the field so as to be able to identify the cases that may be treated with drugs, and know which cases require the approach taken hitherto (such as attempting to reinstate a word's representation).

REFERENCES

Albert, M.L. & Obler, L.K. (1978). *The bilingual brain: Neuropsychological and neurolinguistic aspects of bilingualism.* New York/London: Academic Press.

Allport, D.A. (1985). Distributed memory, modular subsystems and dysphasia. In S. Newman & R. Epstein (eds.), *Current perspectives in dysphasia*, pp. 32-60. Edinburgh: Churchil Livingstone.

Allport, A., Styles, E.A. & Hsieh, S. (1994). Shifting intentional set: exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.). *Attention and performance XV*, pp. 421-452. Cambridge, MA: MIT Press.

Allport, A. & Wylie, G (2000). Task switching, stimulus-response bindings and negative priming. In S. Monsell and J. Driver (Eds.) *Control of cognitive processes. Attention and performance XVIII*, pp. 35-71. Cambridge, MA: MIT Press.

Altenberg, E. & Cairns, H. (1983). The effects of phonotactic constraints in lexical processing in bilingual and monolingual subjects. *Journal of Verbal Learning and Verbal Behavior*, **22**, 174-188.

Anderson, J.R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, **22**, 261-295.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory and Cognition*, **15**, 802-814.

Azuma, R. & Monsell, S. (2000, in preparation). Task switching and compatibility between stimulus attribute and response mode.

Baayen, H., Piepenbrock, R. & van Rijn, H. (1993). *The CELEX lexical database* (CD-ROM). Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.

Bachman, L.F. & Palmer, A. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing*, **6**, 14-29.

Balota, D.A. (1994). Visual word recognition: The journey from features to meaning. In M.A. Gernsbacher (ed.), *Handbook of psycholinguistics*, pp. 303-358. London: Academic Press.

Barron, R.W. & Henderson, L. (1977). The effects of lexical and semantic information on same-different visual comparison of words. *Memory & Cognition*, **5**, 566-579.

Battig, W.F. & Montague, W.E. (1969). Category norms of verbal items in 56 categories A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, **80**(3, Pt. 2), 1-46.

Beardsmore, H.-B. (1980). On the similarities between bilingualism and unilingualism. In P.H. Nelde (ed.), *Languages in contact and conflict*, pp. 11-17. Wiesbaden, Germany:

Franz Steiner Verlag.

Beauvillain, C. & Grainger, J. (1987). Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of Memory and Language*, **26**, 658-672.

Besner, D., Smith, M.C. & MacLeod, C.M. (1990).Visual word recognition: A dissociation of lexical and semantic processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **16**, 862-869.

Bijeljac-Babic, R., Biardeau, A. & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory & Cognition*, **25**, 447-457.

Bock, K. & Levelt, W. (1994). Language production: Grammatical encoding. In M.A. Gernsbacher (ed.), *Handbook of psycholinguistics*, pp. 945-984. London: Academic Press.

Braver, T.S. & Cohen, J.D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In S. Monsell & J. Driver (eds.), *Attention and performance XVIII*, pp. 713-737. Cambridge, MA/London: MIT Press.

Caramazza, A. & Brones, I. (1980). Semantic classification by bilinguals. *Canadian Journal of Psychology*, **34**, 77-81.

Chambers, S.M. (1979). Letter and order information in lexical access. *Journal of Verbal Learning and Verbal Behavior*, **18**, 225-241.

Chumbley, J.I. & Balota, D.A. (1984). A word's meaning affects the decision in lexical decision. *Memory and Cognition*, **12**, 590-606.

Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, **12**, 335-359.

Clyne, M.G. (1980). Triggering and language processing. *Canadian Journal of Psychology*, **34**, 400-406.

Coleman, E.B. (1979). Generalisation effects vs. random effects: Is σ^{2}_{TL} a source of type 1 or type 2 error? *Journal of Verbal Learning and Verbal Behaviour*, **18**, 243-256.

Collins, A.M. & Loftus, E.F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, **82**, 407-428.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, **33A**, 497-505.

Coltheart, M., Davelaar, E., Jonasson, J.T. & Besner, D. (1977). Access to the internal lexicon. In S. Dornič (ed.), *Attention and performance VI*, pp. 535-555. New York: Academic Press.

Dalrymple-Alford, E.C. (1985). Language switching during bilingual reading. *British Journal of Psychology*, **76**, 111-122.

Dell, G.S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, **93**, 283-321.

Dijkstra, T., de Bruijn, E., Schriefers, H. & ten Brinke, S. (2000). More on interlingual homograph recognition: Language intermixing versus explicitness of instructions. *Bilingualism: Language & Cognition*, **3**, 69-78.

Dijkstra, T., Grainger, J. & van Heuven, W.J.B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, **41**, 496-518.

Dijkstra, T. & van Heuven, W.J.B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. Jacobs (eds.), *Localist connectionist approaches to human cognition*, pp. 189-225. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dijkstra, T. & van Heuven, W.J.B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language & Cognition*, **5**, part 3 (in press).

Dijkstra, T., van Heuven, W.J.B. & Grainger, J. (1998). Simulating cross-language competition with the Bilingual Interactive Activation model. *Psychologica Belgica*, **38**, 177-197.

Dijkstra, T., van Jaarsfeld, H. & ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language & Cognition*, **1**, 51-66.

Dijkstra, T., Timmermanns, M. & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language*, **42**, 445-464.

Dillon, R., McCormack, P., Petrusic, W., Cook, G. & Lafleur, L. (1973). Release from proactive interference in compound and coordinate bilinguals. *Bulletin of the Psychonomic Society*, **2**, 293-294.

Dorfman, D. & Glanzer, M. (1988). List composition effects in lexical decision and recognition memory. *Journal of Memory and Language*, **27**, 633-648.

Dornič, S. (1978). The bilingual's performance: Language dominance, stress and individual differences. In D. Gerver & H. Sinaiko (eds.), *Language interpretation and communication*, pp. 259-271. New York: Plenum Press.

Dornič, S. (1979). Information processing in bilinguals: Some selected issues. *Psychological Research*, **40**, 329-348.

Dornič, S. & Wirberg, A. (1983). Note on intralingual and interlingual interference in the Stroop task. *Reports from the Department of Psychology, University of Stockholm*, **611**, 1-7.

Dufour R. & Kroll, J.F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition*, **23**, 166-180.

Durgunoğlu, A.Y. & Roediger III., H.L. (1987). Test differences in accessing bilingual memory. *Journal of Memory and Language*, **26**, 377-391.

Dyer, F.N. (1971). Colour-naming interference in monolinguals and bilinguals. *Journal of Verbal Learning and Verbal Behavior*, **10**, 297-302.

Ellis, A.W. & Young, A.W. (1988). *Human cognitive neuropsychology*. London: Lawrence Erlbaum Associates.

Ervin, S. & Osgood, C. (1954). Second language learning and bilingualism. *Journal of Abnormal and Social Psychology*, **49**, 139-146.

Fabbro, F. & Paradis, M. (1995). Acquired aphasia in a bilingual child. In M. Paradis (ed.), *Aspects of bilingual aphasia*, pp. 67-83. Oxford: Elevier Science.

Fang, S.P., Tzeng, O.J.L. & Alva, L. (1981). Intralingual versus interlingual Stroop effects in two types of writing systems. *Memory & Cognition*, **9**, 609-617.

Farah, M.J., McMullen, P.A. & Meyer, M.M. (1991). Can recognition of living things be selectively impaired? *Neuropsychologia*, **29**, 185-193.

Forster, K.I. (1976). Accessing the mental lexicon. In E.C.J. Walker and R.J. Wales (eds.), *New approaches to language mechanisms*, pp. 257-287. Amsterdam: North-Holland.

Forster, K.I. (1979). Levels of processing and the structure of the language processor. In W.E. Cooper and E. Walker (eds.), *Sentence processing*, pp. 27-85. Hillsdale, NJ: Lawrence Erlbaum Associates.

Forster, K.I. & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **10**, 680-698.

Forster, K.I. & Dickinson, R.G. (1976). More on the language-as-fixed-effect fallacy: Monte Carlo estimates of error rates for F_1, F_2, F' and min F'. *Journal of Verbal Learning and Verbal Behaviour*, **15**, 135-142.

French, R.M. & Ohnesorge, C. (1995). Using non-cognate interlexical homographs to study bilingual memory organisation. In J.D. Moore & J.F. Lehman (eds.), *Proceedings of the 17th annual conference of the Cognitive Science Society*, pp. 31-36. Mahwah, NJ: Lawrence Erlbaum Associates.

Funnel, E. & Allport, A. (1987). Non-linguistic cognition and word meanings: Neuropsychological exploration of common mechanisms. In A. Allport & E. Scheerer (eds.), *Language perception and production*, pp. 367-400. London: Academic Press.

Gerard, L.D. & Scarborough, D.L (1989). Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **15**, 305-315.

Gernsbacher, M.A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness and polysemy. *Journal of Experimental Psychology: General*, **113**, 256-281.

Gernsbacher, M.A. & Faust, M.E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **17**, 245-262.

Glanzer, M. & Ehrenreich, S.L. (1979). Structure and search of the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, **18**, 381-398.

Gollan, T., Forster, K.I. & Frost, R. (1997). Cross-language priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **23**, 1122-1139.

Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior*, **22**, 24-44.

Gordon, B. (1985). Subjective frequency and the lexical decision latency function: Implications for mechanisms of lexical access. *Journal of Memory and Language*, **24**, 631-645.

Grainger, J. (1992). Orthographic neighborhoods and visual word recognition. In R. Frost and L. Katz (eds.), *Orthography, phonology, morphology, and meaning*, pp. 131-146. Amsterdam: Elsevier Science Publishers B.V.

Grainger, J. & Beauvillain, C. (1987). Language blocking and lexical access in bilinguals. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, **39A**, 295-319.

Grainger, J. & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In R.J. Harris (ed.), *Cognitive processing in bilinguals*, pp. 207-220. Amsterdam: Elsevier Science Publishers B.V.

Grainger, J. & Frenck-Mestre, C. (1998). Masked priming by translation equivalents in proficient bilinguals. *Language and Cognitive Processes*, **13**, 601-623.

Grainger, J. & Jacobs, A.M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, **103**, 518-565.

Grainger, J., O'Regan, J.K, Jacobs, A.M. & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & Psychophysics*, **45**, 189-195.

Grainger, J. & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception & Psychophysics*, **47**, 191-198.

Green, D.W. (1986). Control, activation and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language*, **27**, 210-223.

Green, D.W. (1993). Towards a model of L2 comprehension and production. In R. Schreuder & B. Weltens (eds.). *The bilingual lexicon*, pp. 249-277. Amsterdam/Philadelphia: John Benjamins.

Green, D.W. (1995). The control of resources in bilingual speakers. Paper presented to *Tennet VI, Theoretical and Experimental Neuropsychology*, Montreal. Abstract in *Brain & Cognition*, **30** (1996), 30, 264.

Green, D.W. (1998a). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, **1**, 67-81.

Green, D.W. (1998b). Schemas, tags and inhibition. Reply to commentators. *Bilingualism: Language and Cognition*, **1**, 100-104.

de Groot, A.M.B. (1990). The locus of the associative-priming effect in the mental lexicon. In D.A. Balota, G.B. Flores d'Arcais and K. Rayner (eds.), *Comprehension Processes in Reading*. pp. 101-122. Hillsdale, NJ: Lawrence Erlbaum Associates.

de Groot, A.M.B. (1992a). Bilingual lexical representations: A closer look at conceptual representations. In R. Frost & L. Katz (eds.) *Orthography, phonology, morphology and meaning*, pp. 389-412. Amsterdam: Elsevier.

de Groot, A.M.B. (1992b). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **18**, 1001-1018.

de Groot, A.M.B. (1993). Word-type effects in bilingual processing tasks: Support for a mixed representational system. In R. Schreuder & B. Weltens (eds.), *The bilingual lexicon*, pp. 27-51. Amsterdam: John Benjamins.

de Groot, A.M.B. (1998). La répresentation lexico-sémantique et l'access lexical chez le bilingue. *Psychologie Française*, **43**, 297-312.

de Groot, A.M.B., Delmaar, P. & Lupker, S.J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory. *Quarterly Journal of Experimental Psychology*, **53A (2)**, 397-428.

de Groot, A.M.B. & Nas, G.J.L. (1991). Lexical representation of cognates and non-cognates in compound bilinguals. *Journal of Memory and Language,* **30**, 90-123.

Grosjean, F. (1982). *Life with two languages.* Cambridge, Massachusetts: Harvard University Press.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, **36**, 3-15.

Grosjean, F. (1997). Processing mixed language: Issues, findings and models. In A.M.B. de Groot & J.F. Kroll (eds.), *Tutorials in bilingualism, psycholinguistic perspectives,* pp. 225-254. Mahwah, NJ: Lawrence Erlbaum Associates.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language & Cognition*, **1**, 131-149.

Grosjean, F. (2001). The bilingual's language modes. In J. Nicol (ed.), *One mind, two languages: Bilingual language processing*, pp. 1-22. Oxford: Blackwell.

Guttentag, R.E., Haith, M.M., Goodman, G.S. & Hauch, J. (1984). Semantic processing of unattended words by bilinguals: A test of the input-switch mechanism. *Journal of Verbal Learning and Verbal Behavior*, **23**, 178-188.

Haberlandt, K. (1994). Methods in reading research. In M.A. Gernsbacher (ed.), *Handbook of psycholinguistics*, pp. 1-31. London, Academic Press.

Hebb, D.O. (1958). A textbook of psychology. Philadelphia: W.B. Saunders Co.

Henik, A., Friedrich, F. & Kellogg, W. (1983). The dependence of semantic relatedness effects upon prime processing. *Memory & Cognition*, **11**, 366-373.

van Heuven, W.J.B. (2000). Visual word recognition in monolingual and bilingual readers: Experiments and computational modelling. University of Nijmegen, PhD thesis.

van Heuven, W.J.B., Dijkstra, T. & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, **39**, 458-483.

Hino, Y., Lupker, S. & Sears, C.R. (1997). The effects of word association and meaning frequency in a cross-modal lexical decision task: Is the priming due to "semantic" activation? *Canadian Journal of Experimental Psychology*, **17**, 181-194.

Illes, J., Francis, W.S., Desmond, J.E., Gabrieli, J.D.E., Glover, G.H., Poldrack R., Lee, C.J. & Wagner, A.D. (1999). Convergent cortical representation of semantic processing in bilinguals. *Brain and Language*, **70**, 347-363.

Johnson, N.F. (1992). On the role of cohorts or neighbours in visual word recognition. In R. Frost and L. Katz (eds.), *Orthography, phonology, morphology, and meaning*, pp.

147-164. Amsterdam: Elsevier Science Publishers B.V.

Johnson-Laird, P.N. (1987). The mental representation of the meaning of words. *Cognition*, **25**, 189-211.

Just, M.A. & Carpenter, P.A. (1987). *The psychology of reading and language comprehension*. Newton, MA: Allyn and Bacon.

Kellas, G., Ferraro, F.R. & Simpson, G.B. (1988). Lexical ambiguity and the time course of attentional allocation in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 601-609.

Kirsner, K., Brown, H.L., Abrol, S., Chadha, N.N. & Sharma, N.K. (1980). Bilingualism and lexical representation. *Quarterly Journal of Experimental Psychology*, **4**, 585-594.

Kirsner, K., Lalor, E. & Hird, K. (1993). The bilingual lexicon: Exercise, meaning, and morphology. In R. Schreuder & B. Weltens (eds.), *The bilingual lexicon*, pp. 215-248. Amsterdam/Philadelphia: John Benjamins.

Kirsner, K., Smith, M.C., Lockhart, R.S., King, M.L. & Jain, M. (1984). The bilingual lexicon: Language-specific units in an integrated network. *Journal of Verbal Learning and Verbal Behavior*, **23**, 519-539.

Kiyak, H.A. (1982). Interlingual interference in naming colour words. *Journal of Cross-Cultural Psychology*, **13**, 125-135.

Kolers, P.A. (1963). Interlingual word associations. *Journal of Verbal Learning and Verbal Behavior*, **2**, 291-300.

Kolers, P.A. (1966). Reading and talking bilingually. *American Journal of Psychology*, **79**, 357-376.

Kolers, P.A. (1968). Bilingualism and information processing. *Scientific American*, pp. 78-86.

Kroll, J.F. (1993). Accessing conceptual representations for words in a second language. In R. Schreuder & B. Weltens (eds.), *The bilingual lexicon*, pp. 53-81. Amsterdam: John Benjamins.

Kroll, J.F. & Dijkstra, T. (2001). The bilingual lexicon. In R. Kaplan (ed.) *Handbook of applied linguistics*, pp. 301-321. Oxford, UK: Oxford University Press.

Kroll, J.F. & de Groot, A.M.B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A.M.B. de Groot & J.F. Kroll (eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, pp. 169-199. Mahwah, NJ: Lawrence Erlbaum Associates.

Kroll, J.F. & Sholl, A. (1992). Lexical and conceptual memory in fluent and non fluent bilinguals. In R.J. Harris (ed.), *Cognitive Processing in Bilinguals*, pp. 191-204.

Amsterdam: North-Holland.

Kroll, J.F. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, **33**, 149-174.

Kučera, H. & Francis, W.N. (1967). *Computational analysis of present-day American English.* Providence, Rhode Island: Brown University Press.

LaBerge, D. (1981). Unitization and automaticity in perception. In J.H. Flowers (ed.), *Nebraska symposium on motivation: 1980*, pp. 53-71. Lincoln, NE: University of Nebraska Press.

Lambert, W., Havelka, J. & Crosby, C. (1958). The influence of language acquisition contexts on bilingualism. *Journal of Abnormal and Social Psychology*, **56**, 239-244.

Landauer, T.K. & Freedman, J.L. (1968). Information retrieval from long-term memory; category size and recognition time. *Journal of Verbal Language and Verbal Behavior*, **7**, 291-295.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W.J.M., Roelofs, A. & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, **22**, 1-75.

Lima, S.D. & Huntsman, L.A. (1997). Sequential dependencies in the lexical decision task. *Psychological Research,* **60**, 264-269.

Luce, P.A. & Pisoni, D.B. (1998). Recognizing spoken words: The neighbourhood activation model. *Ear and Hearing*, **19**, vol. 1, 1-36.

Lukatela, G. & Turvey, M.T. (1994). Visual lexical access is initially phonological: I. Evidence from associative priming by words, homophones and pseudohomophones. *Journal of experimental Psychology: General*, **123**, 107-128.

Macnamara, J. (1967). The bilingual's linguistic performance: A psychological overview. *Journal of Social Issues*, **23,** 58-77.

Macnamara, J., Krauthammer, M. & Bolgar, M. (1968). Language switching in bilinguals as a function of stimulus and response uncertainty. *Journal of Experimental Psychology*, **78**, 208-215.

Macnamara, J. & Kushnir, S.L. (1971). Linguistic independence of bilinguals: The input-switch. *Journal of Verbal Learning and Verbal Behavior*, **10**, 480-487.

Mägiste, E. (1984). Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. *Journal of Experimental Psychology: Learning, Memory and*

Cognition, **10**, 304-315.

Mason, W.T.M. (1994). Language switching in bilinguals: Evidence from the semantic categorisation task. Unpublished undergraduate project. Cambridge University, UK.

McClelland, J.L. & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception. Part 1: An account of basic findings. *Psychological Review*, **88**, 375-405.

McClelland, J.L. & Rumelhart, D.E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, **114**, 159-188.

McCormack, P.D. (1974). Bilingual linguistic memory: independence or interdependence; two stores or one? In S.T. Carey (ed.), *Bilingualism, biculturalism and education*, pp. 115-128. Edmonton, Canada: University of Alberta.

McCormack, P.D. (1977). Bilingual linguistic memory: The independence-interdependence issue revisited. In P.A. Hornby (ed.), *Bilingualism: Psychological, social and educational implications*, pp. 57-66. New York: Academic Press.

Meara, P. (1994). *LLEX: Lingua vocabulary tests v1.4.* Centre for Applied Language Studies; Swansea, University of Wales.

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **22**, 1-20.

Meiran, N., Chorev, Z. & Sapir, A. (2000). Component processes in task-switching. *Cognitive Psychology*, **41(3)**, 211-253.

Meuter, R.F.I. (1994). *Language switching in naming tasks.* Unpublished D. Phil. thesis, University of Oxford.

Meuter, R.F.I. & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, **40**, 25-40.

Meyer, D.E. & Ruddy, M.G. (1974). Bilingual word-recognition: Organisation and retrieval of alternative lexical codes. Paper presented at the Eastern Psychological Association, Philadelphia.

Monsell, S. (1996). Control of mental processes. In V. Bruce (ed.), *Unsolved mysteries of the mind: Tutorial essays in cognition*, pp. 93-148. Hove, England: Taylor & Francis.

Monsell, S., Doyle, M.C. & Haggard, P.N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, **118**, 43-71.

Monsell, S. & Driver, J. (2000). Banishing the control homunculus. In S. Monsell & J. Driver (eds.), *Attention and performance XVIII*, pp. 3-32. Cambridge, MA: MIT Press.

Monsell, S., Eimer, M., le Pelley, M., Stafford, S. & Yeung, N. (2000, in preparation). Is response activation postponed or prolonged by a task switch? A lateralized readiness potential study.

Monsell, S., Williams, N., Wright, I. & Rogers, R. (1996, in preparation). Costs of switching between colour-naming and word-naming tasks need not be asymmetrical.

Monsell, S., Yeung, N.P. & Azuma, R. (2000). Reconfiguration of task set: Is it easier to switch to the weaker task? *Psychological Research*, **63**, 250-264.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, **76**, 340-354.

Morton, J. (1996). How many routes to reading? In D.W. Green (ed.), *Cognitive science: An introduction*, pp. 148-173. Cambridge, MA: Blackwell Publishers.

Morton, J. & Frith, U. (1995). Causal modelling: A structural approach to developmental psychopathology. In D. Cicchetti & D.J. Cohen (eds.), *Developmental Psychopathology, Volume 1: Theory and Methods*, pp. 357-390. New York, NY: John Wiley and Sons.

Nas, G. (1983). Visual word recognition in bilinguals: Evidence for a cooperation between visual and sound based codes during access to a common lexical store. *Journal of Verbal Learning and Verbal Behavior*, **22**, 526-534.

Neill, W.T. (1997). Episodic retrieval in negative priming and repetition priming. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **23**, 1291-1305.

Neufeld, G. (1976). The bilingual's lexical store. *ITL Review of Applied Linguistics*, **14**, 15-35.

Neumann, E., McCloskey, M.S. & Felio, A.C. (1999). Cross-language positive priming disappears, negative priming does not: Evidence for two sources of selective inhibition. *Memory & Cognition*, **27**, 1051-1063.

Norman, D.A. & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R.J. Davidson, G.E. Schwartz & D. Shapiro (eds.), *Consciousness and self-regulation*, vol. 4, pp. 1-18. New York: Plenum Press.

Obler, L.K. & Albert, M.L. (1978). A monitor system for bilingual language processing. In M. Paradis (ed.), *Aspects of bilingualism*, pp. 156-164. Columbia, SC: Hornbeam.

Page, M. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences*, **23**, 443-512.

Paradis, M. (1977). Bilingualism and aphasia. In H. Whitaker & H.A. Whitaker (eds.), *Studies* *in neurolinguistics*, vol. 3, pp. 65-121. New York: Academic Press.

Paradis, M. (1979). Language and thought in bilinguals. In W. McCormack & H. Izzo (eds.), *The Sixth LACUS Forum*, pp. 420-431. Columbia, SC: Hornbeam Press, Inc.

Paradis, M. (1997). The cognitive neuropsychology of bilingualism. In A.M.B. de Groot & J.F. Kroll (eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, pp. 331-354. Mahwah, NJ: Lawrence Erlbaum Associates.

Paradis, M., Goldblum, M.C. & Abidi, R. (1982). Alternate antagonism with paradoxical translation behaviour in two bilingual aphasic patients. *Brain and Language*, **15**, 55-69.

Pashler, H. & Baylis, G. (1991). Procedural learning: II., Intertrial repetition effects in speeded-choice tasks. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **17**, 33-48.

Pavlenko, A. (1999). New approaches to concepts in bilingual memory. *Bilingualism: Language and Cognition*, **2**, 209-230.

Penfield, W. & Roberts, L. (1959). *Speech and brain-mechanisms*. Princeton: Princeton University Press.

Perani, D., Schnur, T., Tettamanti, M., Gorno-Tempini, M., Cappa, S.F. & Fazio, F. (1999). Word and picture matching: A PET study of semantic category effects. *Neuropsychologia*, **37**, 293-306.

Pexman, P.M. & Lupker, S.J. (1999). Ambiguity and visual word recognition: Can feedback explain both homophone and polysemy effects? *Canadian Journal of Experimental Psychology*, **53**, 323-334.

Pinker, S. (1994). *The language instinct*. London, England: Penguin Books.

Poplack, S. (1980). Sometimes I'll start a sentence in English y termino en Español: Toward a typology of code-switching. *Linguistics*, **18**, 581-618.

Potter, M.C., So, K.-F., von Eckhardt, B. & Feldman, L.B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, **23**, 23-38.

Preston, M.S. & Lambert, W.E. (1969). Interlingual interference in a bilingual version of the Stroop color-word task. *Journal of Verbal Learning and Verbal Behavior*, **8**, 295-301.

Price, C.J., Green, D.W. & von Studnitz, R.E. (1999). A functional imaging study of translation and language switching. *Brain*, **122**, 2221-2235.

Quillian, M.R. (1962). A revised design for an understanding machine. *Mechanical Translation*, **7**, 17-29.

Raaijmakers, J.G.W., Schrijnemakers, J.M.C. & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language*, **41**, 416-426.

Rabbit, P.M.A. (1968). Repetition effects and signal classification strategies in serial choice response tasks. *Quarterly Journal of Experimental Psychology*, **20**, 232-240.

Rayner, K. & Sereno, S.C. (1994). Eye movements in reading. In M.A. Gernsbacher (ed.), *Handbook of psycholinguistics*, pp. 57-81. London: Academic Press.

Rogers, R.D. & Monsell, S. (1995). Cost of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, **124**, 207-231.

Rubinstein, H., Garfield, L. & Millikan, J.A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Language and Verbal Behavior*, **9**, 487-494.

Ruoff, A. (1990). *Häufigkeitswörterbuch gesprochener Sprache - 2., unveränderte, Auflage.* Tübingen: Max Niemeyer Verlag.

Saffran, E.M., Schwartz, M.F. & Marin, O.S.M. (1980). Evidence from aphasia: Isolating the components of a production model. In B. Butterworth (ed.), *Language production*, volume 1, pp. 221-241. London: Academic Press.

Sánchez-Casas, R., Davis, C.W. & García-Albea, J.E. (1992). Bilingual lexical processing: Exploring the cognate/non-cognate distinction. *European Journal of Cognitive Psychology*, **4**, 311-322.

Sartory, G. & Job, R. (1988). The oyster with four legs: A neuropsychological study on the interaction of visual and semantic information. *Cognitive Neuropsychology*, **5**, 105-132.

Scarborough, D.L., Cortese, C. & Scarborough, H.S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 1-17.

Scarborough, D.L., Gerard, L. & Cortese, C. (1984). Independence of lexical access in bilingual word recognition. *Journal of Verbal Learning and Verbal Behavior*, **23**, 84-99.

Schwartz, M.F., Marin, O.S.M. & Saffran, E.M. (1979). Dissociations of language functioning in dementia: A case study. *Brain and Language*, **7**, 277-306.

Seidenberg, M.S. & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, **96**, 523-568.

Shaffer, W.O. & LaBerge, D. (1979). Automatic semantic processing of unattended words. *Journal of Verbal Learning and Verbal Behavior*, **18**, 413-426.

Shafiullah, M. & Monsell, S. (1999). The cost of switching between Kanji and Kana while reading Japanese. *Language and Cognitive Processes*, **14**, 567-607.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge: Cambridge University Press.

Shallice, T. (1994). Multiple levels of control processes. In C. Umiltà & M. Moscovitch (eds.), *Attention and performance XV*, pp. 395-420. Hillsdale, NJ: Erlbaum.

Shanon, B. (1982). Identification and classification of words and drawings in two languages. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, **34A**, 135-152.

Smith, M.C. (1979). Contextual facilitation in a letter search task depends on how the prime is processed. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 239-251.

Smith, M.C. (1991). On the recruitment of semantic information for word fragment completion: Evidence from bilingual priming. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **17**, 234-244.

Smith, M.C. (1997). How do bilinguals access lexical information? In J.F. Kroll & A.M.B. de Groot (eds.), *Tutorials in bilingualism, psycholinguistic perspectives*, pp. 145-168. Mahwah NJ: Lawrence Erlbaum Associates.

Smith, M.C., Theodor, L. & Franklin, P.E. (1983). The relationship between contextual facilitation and depth of processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **9**, 697-712.

Soares, C. & Grosjean, F. (1984). Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory and Cognition*, **12**, 380-386.

Spelke, E.S., Breinlinger, K., Macomber, J. & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, **99**, 605-632.

Sridhar, S.N. & Sridhar, K.K. (1980). The syntax and psycholinguistics of bilingual code mixing. *Canadian Journal of Psychology*, **34**, 407-416.

Stone, G.O. & van Orden, G.C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **19**, 744-774.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, **18**, 643-662.

Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, **18**, 645-659.

Thomas, M.S.C. (1997). *Connectionist networks and knowledge representation: The case of bilingual lexical representations*. Unpublished D. Phil. thesis, University of Oxford.

Thomas, M.S.C. & A. Allport (1995). The dynamic control of bilinguals' lexical representations. Paper presented at the *British Psychological Society, Cognitive Section, Annual Conference*. Bristol, September 1995.

Thomas, M.S.C. & Allport, A. (2000). Language switching costs in bilingual visual word recognition. *Journal of Memory and Language*, **43**, 44-66.

Tissot, R., Mounin, G. & Lhermitte, F. (1973). *L'agrammatisme*. Paris: Dessart.

Tzelgov, J., Henik, A. & Leiser, D. (1990). Controlling Stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **16**, 760-771.

Tzelgov, J., Henik, A., Sneg, R. & Baruch, O. (1996). Unintentional word reading via the phonological route: The Stroop effect with cross-script homophones. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **22**, 336-349.

Votaw, M.C. (1992). A functional view of bilingual lexico-semantic organization. In R.J. Harris (ed.). *Cognitive processing in bilinguals*. Amsterdam: Elsevier Science Publishers.

Warrington, E.K. & Shallice, T. (1984). Category-specific semantic impairments. *Brain*, **107**, 829-853.

Weinreich, U. (1953). *Languages in contact*. New York: Linguistic Circle of New York.

Whittlesea, B.W.A. & Cantwell, A.L. (1987). Enduring influence of the purpose of experiences: Encoding-retrieval interactions in word and pseudoword perception. *Memory & Cognition*, **15**, 465-472.

Wickelgren, W.A. (1969). Context-sensitive coding, associate memory, and serial order in (speech) behavior. *Psychological Review*, **76**, 1-15.

Yeung, N. & Monsell, S. (2000, in preparation). Switching between tasks of unequal familiarity: The role of stimulus-attribute and response-set selection.

APPENDICES

APPENDIX 1: Language Background Questionnaire (English version)

1) Background

Christian Name + Initial of Surname ....
Date of Birth ....
Place of Birth ....
Hand used for writing ....

Places lived in > 1 year
(your age at the time?)

	Country	Duration / Age
1)	....	....
2)	....	....
3)	....	....
4)	....	....
5)	....	....

Prolonged / regular stays abroad

		Location	Duration / Age	Regularity
- where?	1)	....	....	....
- how long usually?	2)	....	....	....
(indicate your respective age)	3)	....	....	....
- once, twice yearly, ...	4)	....	....	....
	5)	....	....	....

Father's home country ....

Mother's home country ....

What do you consider your home country? ....

Why?
(e.g. grew up there, feel more at home in the language, your friends there, ...)

Father's occupation ....

Mother's occupation ....

Your own current occupation ....

2) Spoken Language History

For the languages concerning you, give a) the age when you started the language in question at school, b) as of when you were able to really speak it fluently, c) since when you have been using it continuously (from when until today it has been your main language of communication).

Age at which started to speak German (a;b;c) ....
Age at which started to speak English (a;b;c) ....

Any other languages		Language	Age started it at (a;b)
	1)	....	....
	2)	....	....
	3)	....	....

Father's mother-tongue ....

Mother's mother-tongue ....

While you were growing up: Other languages spoken by	Speaker	Languages
	Father	....
		....
		....
	Mother	....
		....
		....
	....	....
	(other carer)	....

Language spoken to you by (**NB** here and in all following % questions: E + G = 100%)	Speaker	Languages	% Each
	Father	....	....
		....	....
	Mother	....	....
		....	....
	Other carers in childhood?	....	....
		....	....

Language spoken amongst	Speaker	Language(s)	% Each
	Father & Mother	....	....
		....	....
	You & your Brothers and Sisters	....	....
		....	....
		....	....

Language spoken at	Location / Situation	Languages	% Each
	School	....	....
		....	....
	College	....	....
		....	....
	Work	....	....
		....	....
	Home	....	....
		....	....
	on Holiday	....	....

Language spoken by you to	Speaker	Languages	% Each
	your Father	....	....
		....	....
	your Mother	....	....
		....	....
	other Carer(s)	....	....
		....	....
	your Siblings	....	....
		....	....
	other Relatives	....	....
		....	....
	your Partner	....	....
		....	....
	your Friends	....	....
		....	....

Amount of Exposure to TV/radio/cassette tapes/videos/comics, etc.

Time	Regularity	Language(s)	% Each
Childhood	....	....	....
	....	....	....
	....	....	....
Adolescence	....	....	....
	....	....	....
	....	....	....
now	....	....	....
	....	....	....
	....	....	....

Languages spoken by your relatives.

(Estimate how much you were exposed to it, e.g. often, seldom, regularly, irregularly, etc.)

	Speaker	Language(s)	Amount of Exposure
paternal	Grandfather	....	....
		....	....
	Grandmother	....	....
		....	....
	Aunts/Uncles	....	....
		....	....
	Cousins	....	....
		....	....
maternal	Grandfather	....	....
		....	....
	Grandmother	....	....
		....	....
	Aunts/Uncles	....	....
		....	....
	Cousins	....	....
		....	....

In which language do you know more poems? (by heart) ....

3) Written Language History

In which language did you originally learn to read? ....

Age at which you started to read ....

Second language that you learned to read in ....

Other languages you can now read in (fairly fluently) ....
....
....

Language you read in most now

Source	Language	%Each
Books	German	....
	English	....
	....	....
Papers	German	....
	English	....
	....	....
Magazines	German	....
	English	....
	....	....
....	German	....
(other)	English	....
	....	....

In which language did you originally learn to write? ....

Age at which you started to write ....

Second language that you learned to write in ....

Other languages that can write in (fairly fluently) ....
....

Language used most for writing now

Type	Language	Regularity
Letters	....	....
Essays	....	....
....	....	....
(other)		

If you kept / do keep a private diary: In which language would you write / do you write?

4) Spontaneous Use

Language you spontaneously (would) do the following in

'Activity'	Language
thinking to self	....
dreaming	....
mental arithmetic	....
counting	....
swearing	....
writing a shopping list	....

5) Education

Language spoken by the majority of your peers during the following (as applicable for you until now). If varied give order and indication of length in years.

Setting	Language	How long you were there
Nursery	....	....
Kindergarten	....	....
Primary School	....	....
Secondary School	....	....
University	....	....
Work	....	....

Native language of majority of others at

Setting	Language
Nursery	....
Kindergarten	....
Primary School	....
Secondary School	....
University	....
Work	....

Foreign languages learned / studied at

Setting	Language	Length of time studied
School	....	....
University	....	....

Any other educational experience in any other language
(e.g. student exchange, year abroad, ...)

Type of Experience	Language	Duration	Profited from the Experience?
....	....	....	....
....	....	....	....
....	....	....	....

6) Language Weighting

Estimate your ability in German and English for the following:
(NB: together both languages should add up to 100 % for each item!)

Type of Usage	Language	Strength(%)
Reading	German	....
	English	....
Writing	German	....
	English	....
Speaking	German	....
	English	....
Comprehending	German	....
	English	....
Overall	German	....
	English	....

Any idea for the reason of this?

Has it always been so?
(if not - how was it before and why/when did it change?)

If you had to lose either German or English - which would it be? Make a choice thinking of which language is most dear to you, not what the practicalities are of how many people speak what language in the world!

7) Type of Language Usage

Do you ever mix languages?
(estimate regularity; e.g. always, frequently, sometimes, seldom, never)

Conversational Partner	Languages mixed	Regularity of Mixing
Parents	....	....
other Carers	....	....
Siblings	....	....
Relatives	....	....
Partner	....	....
Friends	....	....

Give an example of a situation in which you do this:

Do/did your parents / carers ever mix their languages? (If differing give ôthenö & ônowö)

Conversational Partner	Frequency of Mixing
to Each Other	....
to You	....
to Others in Your Hearing	....

Do you have the impression that your languages (sometimes) influence each other?

Yes / No

What sort of influence?
(e.g. grammar, choice of words, the way they sound, ...)

Language	Influenced by
German	....
	....
	....
English	....
	....
	....

If you can’t think of a particular word in the language you are just speaking in - what do you do?

Conversation Partner	Strategy you employ
other Bilinguals: (same languages as you)	
Unilinguals:	

Do you like talking? (how bad would it be for you to become mute?)

Do you like writing? (how bad would it be for you to become paralysed and unable to write?)

8) Development

When did you first start to speak?
If you know, give the age, otherwise 'early', 'normal', 'late': ....

Did you ever go to a speech & language therapist? Y / N

- what was the reason?

- was it regular? (e.g., once or twice weekly,...)

- for how long?

- if you did have speech therapy: how would you describe your speech & language today?

One Final Thing:

Estimate the extent to which you use the two languages at the moment (together both languages should add up to 100 %, for spoken and written respectively),

	German	English
spoken:	............%	............%
written:	.............%	.............%

Many Thanks!

Fragebogen zur Zweisprachigkeit (German version)

1) Hintergrund

Vorname + Initiale des Nachnamens . ...
Geburtsdatum ....
Geburtsort ....
Welche Hand wird zum Schreiben benutzt? ....

Orte an denen Du länger als 1 Jahr gelebt hast
(Dein Alter zu der Zeit?)

	Land	Dauer/Alter (von-bis)
1)	....	....
2)	....	....
3)	....	....
4)	....	....
5)	....	....

Längere / regelmäßige Auslandsaufenthalte
(ein-, zweimal jährlich, ...)

	Land	Alter	Dauer	Regelmäßigkeit
1)	....	....	....	....
2)	....	....	....	....
3)	....	....	....	....
4)	....	....	....	....
5)	....	....	....	....

Heimatland des Vaters ....

Heimatland der Mutter ....

Was siehst Du als Dein Heimatland an? ....

Weshalb siehst Du es als solches an?
(z.B. dort aufgewachsen, fühle mich in der Sprache mehr zuhause, Freunde dort, ...)

Beschäftigung des Vaters ....

Beschäftigung der Mutter ....

Deine derzeitige Beschäftigung ....

2) Gesprochene Sprache

Für die Sprachen die Dich betreffen gib bitte folgendes an: a) wann Du in der Schule begannst die Sprache zu lernen, b) ab wann Du sie fliessend sprechen konntest, c) seit wann Du sie ununterbrochen sprichst (ab wann bis heute).

Alter in dem Du begannst Deutsch zu sprechen (a;b;c): ....
Alter in dem Du begannst Englisch zu sprechen (a;b;c): ....

Andere Sprachen:

	Sprache	Alter in dem Du sie zu sprechen begannst (a/b für jede)
1)	....	....
2)	....	....
3)	....	....

Muttersprache des Vaters: ....
Muttersprache der Mutter: ....

Andere Sprachen gesprochen von:

Sprecher	Sprache
Vater	....
	....
	....
Mutter	....
	....
	....
....	....
(andere Erzieher)	....

Wie wurdest/wirst Du von anderen angesprochen
(NB immer wenn
% Fragen:
D + E = 100%):

Sprecher	Sprache	ca. % pro Sprache
Vater	....	....
	....	....
Mutter	....	....
	....	....
(andere Erzieher)	....	....
	....	....

Untereinander sprechen wir / sprachen sie am meisten:

Sprecher	Sprache(n)	ca. % pro Sprache
Vater & Mutter	....	....
	....	....
Du & Deine Geschwister	....	....
	....	....

Was wurde/wird wo gesprochen:

Ort / Situation	Sprachen	ca. % pro Sprache
Schule	....	....
	....	....
Universität	....	....
	....	....
Arbeit	....	....
	....	....
zu Hause	....	....
	....	....
Im Urlaub	....	....
	....	....

Wie sprachst/sprichst Du sie an:

Sprecher	Sprachen	ca. % pro Sprache
Deinen Vater	....	....
	....	....
Deine Mutter	....	....
	....	....
andere im Haushalt	....	....
	....	....
Deine Geschwister	....	....
	....	....
andere Verwandte	....	....
	....	....
Dein/e Freund/in	....	....
	....	....
Deine Freunde	....	....
	....	....

Einfluß der Medien (Kassetten, Fernsehen, Radio, Video, Comics, etc.) im Laufe der Zeit:

Alter	Regelmäßigkeit	Sprachen	ca. % pro Sprache
Kindheit	....	....	....
	....	....	....
	....	....	....
frühe Jugend	....	....	....
	....	....	....
	....	....	....
heute	....	....	....
	....	....	....
	....	....	....

Welche Sprachen sprechen Deine Verwandten?

(Schätze wie viel /oft Du damit in Berührung kamst/kommst, z.B. oft, manchmal, selten, gar nicht, regelmäßig [ein-, zweimal jährlich, ...], unregelmäßig [alle Jahre mal, ...], ...)

	Sprecher	Sprache(n)	wie oft damit in Berührung?
väterlicherseits:			
	Großvater	....	....
		....	....
	Großmutter	....	....
		....	....
	Tanten/Onkel	....	....
		....	....
	Vettern/Cousinen	....	....
		....	....
mütterlicherseits:			
	Großvater	....	....
		....	....
	Großmutter	....	....
		....	....
	Tanten/Onkel	....	....
		....	....
	Vettern/Cousinen	....	....
		....	....

In welcher Sprache kannst Du mehr Gedichte auswendig? ....

3) Schriftsprache

In welcher Sprache hast Du ursprünglich Lesen gelernt? ....

Mit wieviel Jahren hast Du Lesen gelernt? ....

Welche Sprache ist die zweite in der Du Lesen lerntest? ....

In welchen anderen Sprachen kannst Du (gut) lesen? ....
....
....

Sprachen in denen Du jetzt am meisten liest:

Material	Sprache	ca.% pro Sprache
Bücher	Deutsch	....
	Englisch	....
	....	....
Zeitung	Deutsch	....
	Englisch	....
	....	....
Zeitschriften	Deutsch	....
	Englisch	....
	....	....
....	Deutsch	....
(anderes)	Englisch	....
	....	....

In welcher Sprache lerntest Du ursprünglich Schreiben? ....

Wann lerntest Du zu schreiben? (Alter) ....

In welcher Sprache lerntest Du als zweites Schreiben? ....

Andere Sprachen in denen Du (gut) schreiben kannst: ....
....

Sprache in der Du z.Z. am meisten schreibst:

Schriftstück	Sprache	Regelmäßigkeit
Briefe	....	....
Aufsätze	....	....
.... (anderes)	....	....

Wenn Du ein Tagebuch führtest - in welcher Sprache tätest Du dies? ....

4) Spontan-Sprache

Sprache in der Du die folgenden Dinge spontan tätest/tust:

‘Tätigkeit’	Sprache
Denken	....
Träumen	....
Kopfrechnen	....
Zählen	....
Fluchen	....
Einkaufsliste schreiben	....

5) Ausbildung

Sprache die von der Mehrzahl der Gleichaltrigen um Dich gesprochen wurde/wird; von damals bis heute - so weit wie zutreffend:
(Falls wechselnd innerhalb einer Kategorie, bitte in Reihenfolge und mit Angabe zur Dauer angeben.)

Ort	Sprache	Wie lange warst Du dort?
Kleinkinder-Hort	....	....
Kindergarten	....	....
Unterstufe	....	....
Oberstufe	....	....
Universität	....	....
Arbeit	....	....

Muttersprache der Mehrzahl der Gleichaltrigen:

Ort	Sprache
Hort	....
Kindergarten	....
Unterstufe	....
Oberstufe	....
Universität	....

Fremdsprachen die Du gelernt hast & für wie lange Du sie lerntest:

Ort	Sprache	Dauer
Schule	....	....
	....	....
Universität	....	....
	....	....

Andere Sprach-Erfahrungen im Rahmen der Ausbildung:
(z.B. Schüleraustausch, Auslandssemester, ...)

Art der Sp-Er.	Sprache	Dauer	davon profitiert?
....	....	....	....
....	....	....	....
....	....	....	....

6) Abwägung der Sprachen gegeneinander

Welche Sprache siehst Du als die stärkere an? (d.h. in welcher bist Du besser)
(Gewichtung in geschätzten % angeben - zusammen gezählt sollen die Sprachen pro Kategorie 100% geben!)

Art der Nutzung	Sprache	Stärke (%)
Lesen	Deutsch	....
	Englisch	....
Schreiben	Deutsch	....
	Englisch	....
Sprechen	Deutsch	....
	Englisch	....
Verstehen	Deutsch	....
	Englisch	....
Insgesamt	Deutsch	....
	Englisch	....

Siehst Du einen Grund weshalb dies so ist?

Ist das schon immer so gewesen?
(Falls nein - was war vorher und wann/warum änderte es sich?)

Wenn Du Deutsch oder Englisch aufgeben müßtest - welches würdest Du aufgeben? Denke hier daran welche Sprache Dir am nähesten ist - nicht an praktische Aspekte (z. B. wieviel Menschen in der Welt welche Sprache sprechen)!

7) Sprachgebrauch

Vermischst Du je Deine Sprachen? Mit wem?
(schätze die Regelmäßigkeit davon, z.B. immer, oft, manchmal, selten, nie, ...)

Gesprächspartner	Sprachen die vermischt werden	Regelmäßigkeit
Eltern	....	....
andere Erzieher (z.B. au-pair)	....	....
Geschwister	....	....
Verwandte	....	....
Freund/in	....	....
Freundeskreis	....	....

Ein Beispiel einer Situation in der Du mischtest:

Vermischen/-ten Deine Eltern jemals ihre Sprachen? Wem gegenüber?

Gesprächspartner	Regelmäßigkeit des Mischens
Eltern untereinander	....
Eltern Dir gegenüber	....
Eltern zu anderen in Deiner Gegenwart	....

Hast Du manchmal den Eindruck, daß Deine beiden Sprachen einander beeinflussen?

Ja / Nein

Welcher Art ist dieser Einfluß?
(z.B. grammatikalisch, in Bezug auf Wortwahl, Wortklang, ...)

Beeinflußte Sprache	Art des Einflusses
Deutsch	....
	....
	....
Englisch	....
	....
	....

Wenn Dir in der Sprache in der Du gerade sprichst ein bestimmtes Wort nicht einfällt, was tust Du dann?

Gesprächspartner	Strategie
andere Zweisprachige (gleiche Sprachen wie Du)	
Einsprachige (wenig oder kein Verständnis Deiner Sprachen)	

Sprichst Du gerne? Wie schlimm wäre es für Dich nicht mehr sprechen zu können?

Schreibst Du gerne? Wie schlimm wäre es für Dich nicht mehr schreiben zu können?

8) Entwicklung

Wann hast Du angefangen zu sprechen?
Alter falls bekannt, sonst 'früh'/'normal'/'spät': ....

Bist Du je bei einer Logopädin gewesen? Ja / Nein

Falls ja:

- was war der Grund hierfür?

- gingst Du regelmäßig?
  (z.B.ein-, zweimal wöchentlich, ...)

- für wie lange war dies?

- falls Du zu einer Logopödin gingst: wie würdest Du Deine Sprache wie sie heute ist beschreiben?

Und zum Schluß:

Schätze wieviel Du die beiden Sprachen momentan gebrauchst (jeweils für gesprochen und geschrieben, sollen beide Sprachen zusammen 100 % ergeben),

	Englisch	Deutsch
gesprochene Sprache:	.............%	.............%
Schriftsprache:	.............%	.............%

Vielen Dank!

APPENDIX 2: Self-Assessment Scale for Languages

(adapted from Bachman & Palmer, 1989)

Please tick the appropriate box - for each question (1-7) give an answer for BOTH German & English:

1) How many different kinds of grammatical mistakes do you make in

	German	English
I make grammatical mistakes in almost everything	☐	☐
Many kinds	☐	☐
Only a few kinds	☐	☐
I almost never make grammatical mistakes	☐	☐

2) How often do you think you don't know enough words in

	German	English
Almost always	☐	☐
Often	☐	☐
Not very often	☐	☐
Never	☐	☐

3) How hard is it for you to put several sentences in a row in

	German	English
Impossible	☐	☐
Very hard	☐	☐
Not very hard	☐	☐
Very easy	☐	☐

4) How hard is it for you to organise a speech or piece of writing containing several ideas in

	German	English
Impossible	☐	☐
Very hard	☐	☐
Not very hard	☐	☐
Very easy	☐	☐

5) How hard is it for you to use different kinds of the two languages with different kinds of people? (e.g. a child, a close friend, a teacher/lecturer/boss)

	German	English
Impossible	☐	☐
Very hard	☐	☐
Not very hard	☐	☐
Very easy	☐	☐

6) When you use the two languages, how hard is it for you to use the same words and sentences that native speakers of the language would use?

	German	English
Impossible	☐	☐
Very hard	☐	☐
Not very hard	☐	☐
Very easy	☐	☐

7) How hard is it for you to use names of well-known English/German people and places in your speaking and writing?

	German	English
Impossible	☐	☐
Very hard	☐	☐
Not very hard	☐	☐
Very easy	☐	☐

APPENDIX 3: Experiment 1a, summary of analytical details (word and nonword data) not mentioned in the body of the text.

Repeated-measures ANOVA. Analyses by subjects, for words: within-subject factors are Trial Type (switch/non-switch), Language (English/German) and Frequency (HF/LF). For nonwords Frequency is replaced by Nonword Type (non-unique/unique). Analyses by items have Language, Frequency and Nonword Type as between-subject factors.

		RT analyses	Error analyses
Word data			
	Frequency	$F_1(1,19) = 83.53$, $p < 0.001$ $F_2(1,296) = 62.87$, $p < 0.001$ min $F'(1,93) = 35.87$, $p < 0.001$	$F_1(1,19) = 269.46$, $p < 0.001$ $F_2(1,307) = 77.25$, $p < 0.001$ min $F'(1,218) = 60.04$, $p < 0.001$
	Language	$F_1(1,19) = 15.50$, $p < 0.001$ $F_2(1,296) = 11.57$, $p < 0.001$ min $F'(1,93) = 6.63$, $p = 0.01$	$F_1 < 1$ $F_2(1,307) = 10.56$, $p < 0.001$ min $F' < 1$
	Frequency x Language	$F_1(1,19) = 17.00$, $p < 0.001$ $F_2(1,296) = 4.40$, $p < 0.05$ min $F'(1,229) = 3.50$, $p = 0.062$	$F_1(1,19) = 21.86$, $p < 0.001$ $F_2(1,307) = 15.69$, $p < 0.001$ min $F'(1,97) = 9.13$, $p < 0.01$
Nonword data			
	Language	$F_1(1,19) = 50.25$, $p < 0.01$ $F_2 < 1$ min $F' < 1$	$F_1(1,19) = 7.25$, $p < 0.025$ $F_2(1,256) = 12.10$, $p < 0.001$ min $F'(1,47) = 4.53$ $p < 0.05$
	Language x Nonword Type	$F_1(1,19) = 10.73$, $p < 0.025$ $F_2(1,246) = 3.19$, n.s. min $F'(1,192) = 2.46$, n.s.	$F_1(1,19) = 5.55$, $p < 0.05$ $F_2(1,256) = 1.60$, n.s. min $F'(1,204) = 1.24$, n.s.

APPENDIX 4: Experiment 1b, summary of analytical details (word and nonword data) not mentioned in the body of the text.

Repeated-measures ANOVA. Analyses by subjects, for words: within-subject factors are Trial Type (switch/non-switch), Language (English/German) and Frequency (HF/LF). For nonwords Frequency is replaced by Nonword Type (non-unique/unique). Analyses by items have Language, Frequency and Nonword Type as between-subject factors.

		RT analyses	Error analyses
Word data			
	Frequency	$F_1(1,18) = 64.52$, $p < 0.001$ $F_2(1,291) = 200.76$, $p < 0.001$ min $F'(1,31) = 48.83$, $p < 0.001$	$F_1(1,18) = 225.47$, $p = 0.001$ $F_2(1,316) = 73.61$, $p < 0.001$ min $F'(1,194) = 55.49$, $p < 0.001$
	Language	$F_1(1,18) = 9.95$, $p < 0.01$ $F_2(1,291) = 28.13$, $p < 0.001$ min $F'(1,33) = 7.35$, $p < 0.025$	$F_1(1,18) = 4.38$, $p = 0.051$ $F_2(1,316) = 6.04$, $p < 0.025$ min $F'(1,52) = 2.54$, $p = 0.12$
	Trial Type x Language	$F_1(1,18) = 6.61$, $p < 0.025$ $F_2(1,291) = 3.01$, $p = 0.081$ min $F'(1,142) = 2.07$, n.s.	$F_1 < 1$ $F_2 < 1$ min $F' < 1$
	Frequency x Language	$F_1(1,18) = 17.43$, $p < 0.001$ $F_2(1,291) = 9.60$, $p < 0.01$ min $F'(1,119) = 6.19$, $p < 0.05$	$F_1(1,18) = 10.48$, $p < 0.01$ $F_2(1,316) = 7.42$, $p < 0.01$ min $F'(1,94) = 4.34$, $p < 0.05$
Nonword data			
	Language	$F_1(1,18) = 3.34$, n.s. $F_2(1,287) = 2.22$, n.s. min $F' < 1$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$
	Language x Nonword Type	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1(1,18) = 8.49$, $p < 0.01$ $F_2(1,316) = 3.03$, $p = 0.083$ min $F'(1,179) = 2.23$, n.s.

APPENDIX 5: Experiment 2, repeated-measures ANOVA on critical HF words (overall analysis).

Subject analyses have the within-subjects factors Task (language-specific / language-general), Language (English / German), Trial Type (switch / non-switch), and the between-subjects factors Condition (100/0 ; 75/25; 50/50) and Task Order (first task / second task). In the item analyses Language is a between-subjects factor, all others are within-subjects.

These data are subsidiary to Experiment 2, providing mainly information on the language background:

As might be expected for fairly balanced bilinguals who lean somewhat towards their L1, RTs were somewhat slower for English HF words (690ms; *SD* = 151) than for German HF words (659ms; *SD* = 132) [$F_1(1,90) = 38.01$, MSE = 4921.64, $p < 0.001$; $F_2(1,94) = 45.18$, MSE = 13277.71, $p < 0.001$; min $F'(1,182) = 20.64$, $p < 0.001$]. There was no difference between the two languages in terms of error rates (English: 1.9%; German: 2.0%) [all Fs < 1].

Overall, there was a significant interaction between trial type and language, such that the cost was greater for switching into English than into German (English: mean RT on switch trials = 718ms, *SD* = 158; on non-switch trials = 662ms, *SD* = 138; German: mean RT on switch trials = 669ms, *SD* = 129; on non-switch trials = 648ms, *SD* = 134) [$F_1(1,90) = 18.97$, MSE = 2911.07, $p < 0.001$; $F_2(1,94) = 20.86$, MSE = 5922.84, $p < 0.001$; min $F'(1,184) = 8.94$, $p < 0.01$]. In terms of errors too, the accuracy difference between switch and non-switch trials was greater for English (2.9% and 0.9% errors respectively) than for German (2.0% and 1.8% errors respectively) [trial type x language: $F_1(1,90) = 8.13$, MSE = 0.02, $p < 0.01$; $F_2(1,94) = 9.88$, MSE = 0.02, $p < 0.01$; min $F'(1,182) = 4.46$, $p < 0.05$]. This finding was unexpected, seeing that there had been no asymmetry of switch-costs in Experiments 1a and 1b (this difference is discussed in Chapter 6).

APPENDIX 5a: Experiment 2, repeated-measures ANOVA on critical HF words (overall analysis: 75/25 vs. 50/50 conditions).

Subject analyses have the within-subjects factors Task (language-specific / language-general), Language (English / German), Trial Type (switch / non-switch), and the between-subjects factors Condition (50/50; 75/25) and Task Order (first task / second task). In the item analyses Language is a between-subjects factor, all others are within-subjects.

These data are subsidiary to Experiment 2, providing mainly information on the language background:

As might be expected for fairly balanced bilinguals who lean somewhat towards their L1, RTs were somewhat slower for English HF words (685ms; *SD* = 154) than for German HF words (648ms; *SD* = 135) [$F_1(1,60) = 38.86$, MSE = 4491.47, $p < 0.001$; $F_2(1,94) = 83.63$, MSE = 7547.33, $p < 0.001$; min $F'(1,113) = 26.53$, $p < 0.001$]. There was no difference between the two languages in terms of error rates (English and German, each 2.0%) [all Fs < 1].

Overall, there was an interaction between trial type and language (significant by subjects, marginal by items, a trend on min F'), such that the cost was greater for switching into English than into German (English: mean RT on switch trials = 712ms, *SD* = 167; on non-switch trials = 659ms, *SD* = 137; German: mean RT on switch trials = 663ms, *SD* = 131; on non-switch trials = 634ms, *SD* = 137) [$F_1(1,60) = 7.73$, MSE = 2115.19, $p < 0.01$; $F_2(1,94) = 3.03$, MSE = 8028.36, $p = 0.085$; min $F'(1,147) = 2.18$, $p = 0.14$]. In terms of errors too, the accuracy difference between switch and non-switch trials was greater for English (2.6% and 0.7% errors respectively) than for German (1.9% and 2.3% errors respectively) [trial type x language: $F_1(1,60) = 7.56$, MSE = 0.01, $p < 0.01$; $F_2(1,94) = 7.34$, MSE = 0.02, $p < 0.01$; min $F'(1,147) = 3.72$, $p = 0.057$]. That is, there seemed, also for the participants in the two mixed conditions, to be an asymmetry of switch-costs.

APPENDIX 6: Experiment 2, repeated-measures ANOVA on nonwords.

Subject analyses with the within-subjects factors Task (language-specific / language-general), Language (English / German), Trial Type (switch / non-switch), Nonword Type (non-unique / unique) and the between-subjects factors Condition (100/0 ; 75/25 ; 50/50) and Task Order (first task / second task). Item analyses have Language and Nonword Type as between-subject factors. Only the relevant effects are cited.

	Overall analyses		Language-specific task only		Language-general task only	
	RT analyses	Error analyses	RT analyses	Error analyses	RT analyses	Error analyses
Trial Type (TT)	$F_1(1,90) = 46.45$, $p < 0.001$ $F_2(1,378) = 67.51$, $p < 0.001$ min $F'(1,231) = 27.52$, $p < 0.001$	$F_1(1,90) = 3.81$, $p = 0.054$ $F_2 < 1$ min $F' < 1$	$F_1(1,90) = 80.53$, $p < 0.001$ $F_2(1,378) = 90.72$, $p < 0.001$ min $F'(1,270) = 42.66$, $p < 0.001$	$F_1(1,90) = 1.64$, n.s. $F_2 < 1$ min $F' < 1$	$F_1(1,90) = 1.24$, n.s. $F_2(1,379) = 3.34$, $p = 0.068$ min $F' < 1$	$F_1(1,90) = 2.03$, n.s. $F_2 < 1$ min $F' < 1$
Nonword Type (NT)	$F_1(1,90) = 8.04$, $p < 0.01$ $F_2(1,378) = 2.26$, $p > 0.1$ min $F'(1,461) = 1.76$, $p > 0.1$	$F_1(1,90) = 2.12$, $p = 0.15$ $F_2(1,380) = 5.57$, $p < 0.025$ min $F'(1,169) = 1.54$, n.s.	$F_1(1,90) = 2.49$, n.s. $F_2(1,378) = 1.16$, n.s. min $F < 1$	$F_1(1,90) = 9.61$, $p < 0.01$ $F_2(1,380) = 8.60$, $p < 0.01$ min $F'(1,313) = 4.54$ $p < 0.05$	$F_1(1,90) = 5.92$, $p < 0.025$ $F_2(1,379) = 2.85$, $p = 0.092$ min $F'(1,427) = 1.92$ n.s.	$F_1 < 1$ $F_2 < 1$ min $F' < 1$
Language (L)	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1(1,90) = 4.40$, $p < 0.05$ $F_2(1,380) = 43.97$, $p < 0.001$ min $F'(1,109) = 4.00$, $p < 0.05$	$F_1(1,90) = 3.05$, $p = 0.084$ $F_2 < 1$ min $F' < 1$	$F_1(1,90) = 8.96$, $p < 0.01$ $F_2(1,380) = 28.62$, $p < 0.001$ min $F'(1,152) = 6.82$ $p < 0.01$	$F_1(1,90) = 5.71$, $p < 0.025$ $F_2(1,379) = 2.19$, $p = 0.14$ min $F'(1,446) = 1.58$ n.s.	$F_1 < 1$ $F_2(1,380) = 14.70$, $p < 0.001$ min $F' < 1$
Task Order (O)	$F_1(1,90) = 15.20$, $p < 0.001$ $F_2(1,378) = 799.72$, $p < 0.001$ min $F'(1,94) = 14.92$, $p < 0.001$	$F_1(1,90) = 23.58$, $p < 0.001$ $F_2(1,380) = 89.98$, $p < 0.001$ min $F'(1,141) = 18.68$, $p < 0.001$	$F_1(1,90) = 5.74$, $p < 0.025$ $F_2(1,378) = 656.13$, $p < 0.001$ min $F'(1,92) = 5.69$, $p < 0.025$	$F_1(1,90) = 5.85$, $p < 0.025$ $F_2(1,380) = 52.49$, $p < 0.001$ min $F'(1,111) = 5.26$ $p < 0.025$	$F_1(1,90) = 6.87$, $p = 0.01$ $F_2(1,379) = 358.89$, $p < 0.001$ min $F'(1,93) = 6.74$, $p < 0.025$	$F_1(1,90) = 16.42$, $p < 0.001$ $F_2(1,380) = 45.18$, $p < 0.001$ min $F'(1,162) = 12.04$, $p < 0.001$
Condition (C)	$F_1 < 1$ $F_2(2,756) = 23.59$, $p < 0.001$ min $F' < 1$	$F_1 < 1$ $F_2(2,760) = 3.43$, $p < 0.05$ min $F' < 1$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1 < 1$ $F_2(2,758) = 42.92$, $p < 0.001$ min $F' < 1$	$F_1 < 1$ $F_2(2,760) = 7.76$, $p < 0.001$ min $F' < 1$
Task (T)	$F_1(1,90) = 1.08$, n.s. $F_2(1,378) = 106.21$, $p < 0.001$ min $F'(1,92) = 1.07$, n.s.	$F_1(1,90) = 4.18$, $p < 0.05$ $F_2(1,380) = 10.95$, $p = 0.001$ min $F'(1,166) = 3.03$, $p = 0.084$				
T x TT	$F_1(1,90) = 40.39$, $p < 0.001$ $F_2(1,378) = 31.18$, $p < 0.001$ min$F'(1,339) = 17.60$, $p < 0.001$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$				
T x NT	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1(1,90) = 6.29$, $p < 0.025$ $F_2(1,380) = 3.04$, $p = 0.082$ min $F'(1,435) = 2.05$, n.s.				
T x C	$F_1 < 1$ $F_2(2,756) = 23.17$, $p < 0.001$ min $F' < 1$	$F_1 < 1$ $F_2(2,760) = 5.49$, $p < 0.01$ min $F' < 1$				
T x O	$F_1 < 1$ $F_2(1,378) = 54.76$, $p < 0.001$ min $F' < 1$	$F_1(1,90) = 1.90$, n.s. $F_2(1,380) = 1.52$, n.s. min $F' < 1$				

continued overleaf

continued from previous page

	Overall analyses		Language-specific task only		Language-general task only	
	RT analyses	Error analyses	RT analyses	Error analyses	RT analyses	Error analyses
TT x NT	$F_1(1,90) = 26.29$, $p < 0.001$ $F_2(1,378) = 22.04$, $p < 0.001$ min $F'(1,323) = 11.99$, $p < 0.001$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1(1,90) = 7.27$, $p < 0.01$ $F_2(1,378) = 5.75$, $p < 0.025$ min $F'(1,332) = 3.21$, $p = 0.074$	$F_1(1,90) = 1.46$, n.s. $F_2 < 1$ min $F' < 1$	$F_1(1,90) = 19.06$, $p < 0.001$ $F_2(1,379) = 18.13$, $p < 0.001$ min $F'(1,297) = 9.29$ $p < 0.01$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$
TT x L	$F_1(1,90) = 4.67$, $p < 0.05$ $F_2(1,378) = 2.66$, $p = 0.10$ min $F'(1,384) = 1.69$, n.s.	$F_1 < 1$ $F_2(1,380) = 4.43$, $p < 0.05$ min $F' < 1$	$F_1(1,90) = 5.13$, $p < 0.05$ $F_2(1,378) = 4.01$, $p < 0.05$ min $F'(1,334) = 2.25$, n.s.	$F_1 < 1$ $F_2(1,380) = 3.62$, $p = 0.058$ min $F' < 1$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1 < 1$ $F_2(1,380) = 1.24$, n.s. min $F' < 1$
TT x C	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1(2,90) = 1.03$, n.s. $F_2(2,760) = 4.49$, $p < 0.025$ min $F' < 1$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1(2,90) = 3.49$, $p < 0.05$ $F_2(2,760) = 3.01$, $p = 0.05$ min$F'(2,352) = 1.62$, n.s.	$F_1(2,90) = 2.44$, $p = 0.093$ $F_2(2,758) = 1.22$, n.s. min $F' < 1$	$F_1(2,90) = 1.02$, n.s. $F_2(2,760) = 2.30$, n.s. min $F' < 1$
TT x O	$F_1(1,90) = 1.94$, n.s. $F_2(1,378) = 1.39$, n.s. min $F' < 1$	$F_1(1,90) = 1.97$, n.s. $F_2(1,380) = 4.29$, $p < 0.05$ min $F' < 1$	$F_1 < 1$ $F_2(1,378) = 1.99$, n.s. min $F' < 1$	$F_1(1,90) = 4.59$, $p < 0.05$ $F_2(1,380) = 3.86$, $p = 0.05$ min$F'(1,310) = 2.10$, n.s.	$F_1(1,90) = 2.99$, $p = 0.087$ $F_2 < 1$ min $F' < 1$	$F_1 < 1$ $F_2(1,380) = 1.20$, n.s. min $F' < 1$
NT x L	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1(1,90) = 39.51$, $p < 0.001$ $F_2(1,380) = 33.76$, $p < 0.001$ min $F'(1,320) = 18.20$, $p < 0.001$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1(1,90) = 7.94$, $p < 0.01$ $F_2(1,380) = 4.92$, $p < 0.05$ min $F'(1,376) = 3.04$ $p = 0.082$	$F_1(1,90) = 1.23$, n.s. $F_2 < 1$ min $F' < 1$	$F_1(1,90) = 36.00$, $p < 0.001$ $F_2(1,380) = 33.18$, $p < 0.001$ min $F'(1,306) = 17.27$, $p < 0.001$
NT x C	$F_1(2,90) = 8.97$, $p < 0.001$ $F_2(2,756) = 15.32$, $p < 0.001$ min $F'(2,217) = 5.66$, $p < 0.01$	$F_1(2,90) = 1.00$, n.s. $F_2 < 1$ min $F' < 1$	$F_1(2,90) = 2.28$, n.s. $F_2(2,756) = 9.11$, $p < 0.001$ min $F'(2,141) = 1.82$, n.s.	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1(2,90) = 7.77$, $p = 0.001$ $F_2(2,758) = 9.85$, $p < 0.001$ min $F'(2,268) = 4.34$ $p < 0.025$	$F_1(2,90) = 2.20$, n.s. $F_2(2,760) = 1.67$, n.s. min $F' < 1$
NT x O	$F_1(1,90) = 5.53$, $p < 0.025$ $F_2(1,378) = 10.30$, $p = 0.001$ min $F'(1,199) = 3.60$, $p = 0.059$	$F_1(1,90) = 1.93$, n.s. $F_2(1,380) = 1.52$, n.s. min $F' < 1$	$F_1(1,90) = 7.72$, $p < 0.01$ $F_2(1,378) = 7.84$, $p < 0.01$ min $F'(1,288) = 3.89$, $p < 0.05$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1 < 1$ $F_2(1,379) = 5.64$, $p < 0.025$ min $F' < 1$	$F_1(1,90) = 1.48$, n.s. $F_2 < 1$ min $F' < 1$
L x C	$F_1 < 1$ $F_2(2,756) = 1.72$, n.s. min $F' < 1$	$F_1 < 1$ $F_2(2,760) = 2.44$, $p = 0.088$ min $F' < 1$	$F_1 < 1$ $F_2(2,756) = 2.67$, $p = 0.070$ min $F' < 1$	$F_1 < 1$ $F_2(2,760) = 1.02$, n.s. min $F' < 1$	$F_1 < 1$ $F_2(1,758) = 1.27$, n.s. min $F' < 1$	$F_1(2,90) = 2.36$, n.s. $F_2(2,760) = 3.60$, $p < 0.05$ min $F'(2,237) = 1.43$, n.s.
L x O	$F_1(1,90) = 2.56$, n.s. $F_2(1,378) = 8.26$, $p < 0.01$ min $F'(1,150) = 1.95$, n.s.	$F_1(1,90) = 5.15$, $p < 0.05$ $F_2(1,380) = 22.43$, $p < 0.001$ min $F'(1,150) = 4.19$, $p < 0.05$	$F_1(1,90) = 1.62$, n.s. $F_2(1,378) = 8.30$, $p < 0.01$ min $F'(1,128) = 1.36$, n.s.	$F_1(1,90) = 4.36$, $p < 0.05$ $F_2(1,380) = 16.28$, $p < 0.001$ min $F'(1,142) = 3.44$ $p = 0.066$	$F_1(1,90) = 1.15$, n.s. $F_2(1,379) = 2.98$, $p = 0.085$ min $F' < 1$	$F_1(1,90) = 1.39$, n.s. $F_2(1,380) = 8.72$, $p < 0.01$ min $F'(1,120) = 1.20$, n.s.
C x O	$F_1 < 1$ $F_2(2,756) = 20.02$, $p < 0.001$ min $F' < 1$	$F_1 < 1$ $F_2(2,760) = 7.53$, $p = 0.001$ min $F' < 1$	$F_1 < 1$ $F_2(2,756) = 34.02$, $p < 0.001$ min $F' < 1$	$F_1 < 1$ $F_2(2,760) = 5.36$, $p < 0.01$ min $F' < 1$	$F_1 < 1$ $F_2(2,758) = 6.18$, $p < 0.01$ min $F' < 1$	$F_1 < 1$ $F_2(2,760) = 3.05$, $p < 0.05$ min $F' < 1$
NT x TT x O	$F_1(1,90) = 4.24$, $p < 0.05$ $F_2(1,378) = 2.83$, $p = 0.093$ min$F'(1,357) = 1.70$, n.s.	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1(1,90) = 4.24$, $p < 0.05$ $F_2(1,379) = 2.55$, $p = 0.11$ min$F'(1,384) = 1.59$, n.s.	$F_1 < 1$ $F_2 < 1$ min $F' < 1$
TT x L x T	$F_1(1,90) = 2.00$, $p = 0.16$ $F_2(1,378) = 1.60$, n.s. min$F' < 1$	$F_1(1,90) = 1.43$, n.s. $F_2 < 1$ min $F' < 1$				

APPENDIX 7: Experiment 3, repeated-measures ANOVA on (mean correct) RT; for the full and reduced data sets on both languages; for each language.

By subjects, within-subjects effects are Part (part 1 / part 2), Language (English/German), Language-switching (language-switch/language-non-switch), Animacy (animate[1]/inanimate[2]) and Response Type (response-switch/response-repetition). By items, Language and Animacy are between-subjects factors. Note, for space reasons not all interactions are displayed here. Any main effects or interactions not mentioned in this table were n.s. (either F < 1 $p > 0.1$*).*

	RT (full set, E vs. G)	RT (reduced set, E vs. G)[3]	RT (full set, G only)	RT (full set, E only)
Part (P)	$F_1(1,23) = 25.44$ $p < 0.001$ $F_2(1,126) = 117.91$ $p < 0.001$ minF'(1,34) = 20.93, $p < 0.001$	$F_1(1,23) = 35.70$ $p < 0.001$	$F_1(1,23) = 17.61$ $p < 0.001$ $F_2(1,126) = 48.62$ $p < 0.001$ minF'(1,42) = 12.93, $p < 0.001$	$F_1(1,23) = 19.39$ $p < 0.001$ $F_2(1,126) = 54.87$ $p < 0.001$ minF'(1,41) = 14.33, $p < 0.001$
Language (L)	$F_1(1,23) = 35.70$ $p < 0.001$ $F_2(1,126) = 1.29$ $p > 0.2$ minF'(1,134) = 1.25, $p > 0.2$	$F_1(1,23) = 12.23$ $p < 0.01$		
Language-Switching (L-S)	$F_1(1,23) = 6.01$ $p < 0.025$ $F_2(1,126) = 14.34$ $p = 0.001$ minF'(1,45) = 4.24, $p < 0.05$	$F_1(1,23) = 5.90$ $p < 0.05$	$F_1 < 1$ $F_2(1,126) = 3.75$ $p = 0.055$ minF' < 1	$F_1 < 1$ $F_2(1,126) = 3.70$ $p = 0.064$ minF' < 1
P x L-S	$F_1(1,23) = 3.44$ $p = 0.077$ $F_2 < 1$ minF' < 1	$F_1(1,23) = 4.05$ $p = 0.059$	$F_1 < 1$ $F_2 < 1$ minF' < 1	$F_1 < 1$ $F_2 < 1$ minF' < 1
Animacy (A)	$F_1(1,23) = 35.64$ $p < 0.001$ $F_2 < 1$ minF' < 1	$F_1(1,23) = 30.85$ $p < 0.001$	$F_1(1,23) = 65.08$ $p < 0.001$ $F_2(1,126) = 1.16$ $p > 0.2$ minF'(1,130) = 1.14, $p > 0.2$	$F_1(1,23) = 5.29$ $p < 0.05$ $F_2 < 1$ minF' < 1
P x A	$F_1(1,23) = 6.96$ $p < 0.025$ $F_2 < 1$ minF' < 1	$F_1(1,23) = 3.97$ $p = 0.059$	$F_1(1,23) = 2.80$ $p > 0.1$ $F_2 < 1$ minF' < 1	$F_1(1,23) = 4.46$ $p < 0.05$ $F_2 < 1$ minF' < 1
Response Type (R)	$F_1(1,23) = 19.13$ $p < 0.001$ $F_2(1,126) = 8.94$ $p < 0.01$ minF'(1,124) = 6.09, $p < 0.025$	$F_1(1,23) = 2.06$ $p = 0.16$	$F_1(1,23) = 10.68$ $p < 0.01$ $F_2(1,126) = 5.76$ $p < 0.025$ minF'(1,115) = 3.74, $p = 0.055$	$F_1(1,23) = 15.92$ $p = 0.001$ $F_2(1,126) = 3.95$ $p < 0.05$ minF'(1,147) = 3.16, $p = 0.077$
P x R	$F_1(1,23) = 5.25$ $p < 0.05$ $F_2(1,126) = 6.02$ $p < 0.025$ minF'(1,71) = 2.80, $p = 0.098$	$F_1(1,23) = 3.08$ $p = 0.093$	$F_1(1,23) = 1.35$ $p > 0.2$ $F_2(1,126) = 1.21$ $p > 0.2$ minF' < 1	$F_1(1,23) = 2.76$ $p > 0.1$ $F_2(1,126) = 5.34$ $p < 0.025$ minF'(1,51) = 1.82, $p > 0.1$
L x A	$F_1(1,23) = 33.31$ $p < 0.001$ $F_2(1,126) = 1.08$ $p > 0.3$ minF'(1,134) = 1.05, $p > 0.3$	$F_1(1,23) = 6.30$ $p < 0.025$		
L-S x R	$F_1(1,23) = 24.13$ $p < 0.001$ $F_2(1,126) = 15.70$ $p < 0.001$ minF'(1,103) = 9.51, $p < 0.01$	$F_1(1,23) = 5.98$ $p < 0.025$	$F_1(1,23) = 5.81$ $p < 0.025$ $F_2(1,126) = 4.80$ $p < 0.05$ minF'(1,89) = 2.63, $p = 0.11$	$F_1(1,23) = 20.79$ $p < 0.001$ $F_2(1,126) = 12.44$ $p = 0.001$ minF'(1,109) = 7.78, $p < 0.01$
P x Ls x R	$F_1(1,23) = 14.30$ $p = 0.001$ $F_2(1,126) = 7.25$ $p < 0.01$ minF'(1,119) = 4.81, $p < 0.05$	$F_1(1,23) = 4.35$ $p = 0.061$	$F_1(1,23) = 5.17$ $p < 0.05$ $F_2(1,126) = 4.71$ $p < 0.05$ minF'(1,83) = 2.47, $p = 0.12$	$F_1(1,23) = 15.39$ $p = 0.001$ $F_2(1,126) = 3.00$ $p = 0.086$ minF'(1,149) = 2.51, $p = 0.12$

[1] 'animate' = 'living'; [2] 'inanimate' = 'non-living'

[3] Item analyses were not possible for the reduced set, due to the random elimination of items.

APPENDIX 8: Experiments 4a and 4b, additional group comparisons.

Mixed-factor ANOVAs (on the median correct RT data) of the IH interference effect. By subjects, Item Type (IH/control) and Phase (1/2) are within-subjects factors, Condition (Informed+G, Informed-G, Unilingual, Informed, Delayed-informed, Delayed-informed$_{P2}$) is a between-subjects factor. By items, Item Type is between-subjects factor.

	Experiment 4a			**Experiment 4b**	
	Informed+G vs. Informed-G	Informed+G vs. Unilingual	Informed-G vs. Unilingual	**Informed vs. Delayed-informed**	Delayed-informed$_{P2}$ vs. Unilingual
Item Type (IT)	$F_1(1,30) = 29.18$, $p < 0.001$ $F_2(1,76) = 33.32$, $p < 0.001$ min $F'(1,81) = 15.56$, $p < 0.001$	$F_1(1,30) = 23.84$, $p < 0.001$ $F_2(1,76) = 34.71$, $p < 0.001$ min $F'(1,48) = 14.13$, $p < 0.001$	$F_1(1,30) = 40.21$, $p < 0.001$ $F_2(1,76) = 12.60$, $p = 0.001$ min $F'(1,27) = 9.59$, $p < 0.001$	$F_1(1,62) = 70.69$, $p < 0.001$ $F_2(1,76) = 34.84$, $p < 0.001$ min $F'(1,131) = 23.34$, $p < 0.001$	$F_1(1,30) = 29.72$, $p < 0.001$ $F_2(1.76) = 31.86$, $p < 0.001$ min $F'(1,46) = 15.38$, $p < 0.001$
Phase (P)	$F_1(1,30) = 22.06$, $p < 0.001$ $F_2(1,76) = 9.75$, $p < 0.01$ min $F'(1,106) = 6.76$, $p < 0.025$	$F_1(1,30) = 21.72$, $p < 0.001$ $F_2(1,76) = 7.97$, $p < 0.01$ min $F'(1,106) = 5.83$, $p < 0.025$	$F_1(1,30) = 8.52$, $p < 0.01$ $F_2(1,76) = 4.44$, $p < 0.05$ min $F'(1,104) = 2.92$, $p > 0.05$	$F_1(1,62) = 1.93$, n.s. $F_2 < 1$ min $F' < 1$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$
Condition / Group (C/G)	$F_1(1,30) = 4.12$, $p = 0.051$ $F_2(1,76) = 11.70$, $p < 0.001$ min $F'(1,52) = 3.05$, $p = 0.087$	$F_1(1,30) = 12.34$, $p = 0.001$ $F_2(1,76) = 139.89$, $p < 0.001$ min $F'(1,35) = 11.34$, $p < 0.001$	$F_1(1,30) = 18.71$, $p < 0.001$ $F_2(1,76) = 81.53$, $p < 0.001$ min $F'(1,44) = 15.22$, $p < 0.001$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1(1,30) = 13.86$, $p = 0.001$ $F_2(1,76) = 114.17$, $p < 0.001$ min $F'(1,38) = 12.36$, $p < 0.001$
IT x P	$F_1(1,30) = 9.45$, $p < 0.01$ $F_2(1,76) = 2.10$, $p > 0.1$ min $F'(1,100) = 1.72$, n.s.	$F_1(1,30) = 12.42$, $p < 0.001$ $F_2(1,76) = 2.81$, $p = 0.098$ min $F'(1,101) = 2.29$, $p = 0.13$	$F_1 < 1$ $F_2(1,76) = 1.11$, n.s. min $F' < 1$	$F_1 < 1$ $F_2 < 1$ min $F' < 1$	$F_1(1,30) = 1.14$, n.s. $F_2 < 1$ min $F' < 1$
IT x C/G	$F_1(1,30) = 8.69$, $p < 0.01$ $F_2(1,76) = 5.67$, $p < 0.025$ min $F'(1,100) = 3.43$, $p = 0.067$	$F_1(1,30) = 13.73$, $p < 0.001$ $F_2(1,76) = 15.02$, $p < 0.001$ min $F'(1,83) = 7.17$, $p < 0.025$	$F_1(1,30) = 5.33$, $p < 0.05$ $F_2(1,76) = 1.43$, n.s. min $F'(1,104) = 1.13$, n.s.	$F_1(1,62) = 6.33$, $p < 0.025$ $F_2(1,76) = 4.07$, $p < 0.05$ min $F'(1,135) = 2.48$, $p = 0.12$	$F_1(1,30) = 8.98$, $p < 0.01$ $F_2(1,76) = 10.37$, $p < 0.001$ min $F'(1,81) = 4.81$, $p < 0.05$
P x C/G	$F_1(1,30) = 11.17$, $p < 0.01$ $F_2(1,76) = 2.85$, $p = 0.095$ min $F'(1,103) = 2.27$, $p = 0.14$	$F_1(1,30) = 13.85$, $p = 0.001$ $F_2(1,76) = 4.55$, $p < 0.05$ min $F'(1,106) = 3.42$, $p = 0.067$	$F_1 < 1$ $F_2 <1$ min $F' < 1$	$F_1 < 1$ $F_2 <1$ min $F' < 1$	$F_1(1,30) = 1.06$, n.s. $F_2 < 1$ min $F' < 1$
IT x P x C/G	$F_1(1,30) = 12.91$, $p = 0.001$ $F_2(1,76) = 4.63$, $p < 0.05$ min $F'(1,106) = 3.41$, $p = 0.067$	$F_1(1,30) = 15.39$, $p < 0.001$ $F_2(1,76) = 4.14$, $p < 0.05$ min $F'(1,103) = 3.26$, $p = 0.073$	$F_1 < 1$ $F_2 <1$ min $F' < 1$	$F_1 < 1$ $F_2 <1$ min $F' < 1$	$F_1 < 1$ $F_2 <1$ min $F' < 1$

APPENDIX 9: Experiment 2, stimulus list (critical HF stimuli).

Kučera & Francis (1967), Ruoff (1990) and CELEX (Baayen, Piepenbrock & van Rijn, 1993) frequency values, letter length and syllable length for each word serving as a 'critical HF word'.

			Word Frequency			Word Length	
Set	Language	Word	Kučera & Francis	Ruoff	CELEX	Letters	Syllables
A	English	ARTIST	57		39	6	2
		BACK	967		984	4	1
		BED	127		244	3	1
		BRAIN	45		68	5	1
		DUST	70		42	4	1
		EVIL	72		27	4	2
		FIRE	187		146	4	1
		FOREST	66		68	6	2
		IRON	43		68	4	2
		LAKE	54		40	4	1
		LEG	58		63	3	1
		LEVEL	213		180	5	2
		NECK	81		72	4	1
		NOSE	60		73	4	1
		PAPER	157		174	5	2
		POPE	40		13	4	1
		POWER	342		330	5	2
		SENATE	62		15	6	2
		SONG	70		33	4	1
		STAFF	113		117	5	1
		WALK	100		43	4	1
		WATER	442		432	5	2
		WIFE	228		211	4	1
		WIRE	42		35	4	1
	German	ABEND		176	113	5	2
		BERG		130	36	4	1
		BODEN		20	143	5	2
		BRETT		20	3	5	1
		BROT		103	28	4	1
		BUCH		47	99	4	1
		DRECK		34	8	5	1
		ESSEN		78	91	5	2
		FELD		180	58	4	1
		FENSTER		52	75	7	2
		GELD		224	181	4	1
		HAKEN		24	10	5	2
		KARTE		35	19	5	2
		KORB		26	19	4	1
		LEBEN		63	457	5	2
		MONAT		90	73	5	2
		SOMMER		185	63	6	2
		SONNE		22	90	5	2
		STEIN		68	32	5	1
		STUNDE		236	137	6	2
		SUPPE		26	7	5	2
		TOR		23	29	3	1
		VATER		593	191	5	2
		WELT		33	979	4	1

continued overleaf

continued from previous page

			Word Frequency			Word Length	
Set	Language	Word	Kučera & Francis	Ruoff	CELEX	Letters	Syllables
B	English	BAG	42		61	3	1
		CAR	274		276	3	1
		CHEST	53		43	5	1
		DANCE	90		30	5	1
		DATE	119		55	4	1
		DIRT	43		20	4	1
		ENTER	78		12	5	2
		ESTATE	51		37	6	2
		FACE	371		401	4	1
		FRAME	74		25	5	1
		HOME	355		295	4	1
		HORSES	68		48	6	2
		IMAGE	103		74	5	2
		ITEM	54		20	4	2
		POCKET	43		56	6	2
		PRISON	42		69	6	2
		RIVER	165		108	5	2
		SALT	46		35	4	1
		STEP	131		65	4	1
		STONE	58		41	5	1
		SUN	112		150	3	1
		TIME	1599		1789	4	1
		TRAVEL	61		29	6	2
		WOMAN	224		338	5	2
	German	AST		22	3	3	1
		BERUF		50	61	5	2
		BETT		121	80	4	1
		BILD		19	226	4	1
		BLUME		20	3	5	2
		BLUT		134	40	4	1
		DORF		56	53	4	1
		ECKE		21	-	4	2
		FARBE		53	34	5	2
		FIRMA		27	79	5	2
		GRUND		29	239	5	1
		KINO		20	19	4	2
		MUND		3	53	4	1
		PUNKT		22	110	5	1
		SPECK		22	7	5	1
		STANGE		45	5	6	2
		STIFT		1	2	5	1
		WAGEN		141	135	5	2
		WALD		430	51	4	1
		WASSER		140	81	6	2
		WERT		60	95	4	1
		WIESE		73	11	5	2
		WORT		22	124	4	1
		WURST		33	9	5	1

APPENDIX 10: Experiment 3, stimulus list.

B&M = Battig & Montague (1969) word typicality norms; K&F = Kučera & Francis (1967) word frequency norms for English; CELEX (Baayen, Piepenbrock & van Rijn, 1993) = word frequency norms for English [E] and German [G]; Ruoff = (1990) word frequency norms for German; subjective animacy ratings given by English raters (E) and German raters (G) respectively.

English name	B&M	Mean animacy rating (E)	K&F	CELEX (E)	letters / syllables	German Name	Mean animacy rating (G)	Ruoff	CELEX (G)	letters / syllables
Animate: fruits										
Apple	429	6.3	9	18	5 / 2	Apfel	5.6	27	6	5 / 2
Pear	326	6.3	6	2	4 / 1	Birne	5.8	5	0	5 / 2
Grape	247	6.4	3	2	5 / 1	Traube	5.7	14	1	6 / 2
Plum	167	6.5	1	3	4 / 1	Pflaume	5.9	2	0	7 / 2
Berry	11	6.2	9	2	5 / 2	Beere	5.5	19	0	5 / 2
Fig	16	6.0	12	4	3 / 1	Feige	5.8	0	2	5 / 2
Blueberry	21	6.2	1	2	9 / 2	Blaubeere	5.6	0	1	9 / 3
Raisin	16	6.1	1	1	6 / 2	Rosine	5.5	0	1	6 / 3
Animate: vegetables										
Carrot	316	6.2	1	3	6 / 2	Möhre	5.3	0	0	5 / 2
Bean	237	5.8	5	4	4 / 1	Bohne	5.5	5	0	5 / 2
Potato	224	6.5	15	11	6 / 3	Kartoffel	5.4	114	3	9 / 3
Spinach	163	6.7	2	4	7 / 2	Spinat	5.2	2	3	6 / 2
Beets	63	6.0	2	1	5 / 1	Rüben	5.7	57	6	5 / 2
Onions	47	6.6	15	9	6 / 2	Zwiebeln	5.6	1	4	8 / 2
Kale	20	5.4	1	2	4 / 1	Kohl	5.6	2	12	4 / 1
Peppers	13	6.1	13	3	7 / 2	Paprika	5.5	1	1	7 / 3
Animate: flowers										
Tulip	209	6.8	4	1	5 / 2	Tulpe	6.3	0	0	5 / 2
Orchid	135	7.0	3	2	6 / 2	Orchidee	6.5	0	1	8 / 3
Pansy	108	6.7	-	2	5 / 2	Stiefmütt.	6.3	0	2	13 / 4
Daisy	176	6.9	-	7	5 / 2	Gänseblume	6.8	0	0	10 / 4
Poppy	22	6.6	2	2	5 / 2	Mohn	6.3	3	1	4 / 1
Marigold	17	6.5	1	1	8 / 3	Ringelbl.	6.4	0	0	11 / 4
Lilac	45	6.3	4	4	5 / 2	Flieder	6.2	0	1	7 / 2
Sunflower	36	6.9	1	1	9 / 3	Sonnenbl.	6.3	0	0	11 / 4
Animate: trees										
Oak	394	7.1	15	14	3 / 1	Eiche	7.0	2	3	5 / 2
Maple	314	6.6	7	3	5 / 2	Ahorn	6.1	0	1	5 / 2
Elm	210	6.7	3	7	3 / 1	Ulme	6.2	0	0	4 / 2
Birch	134	6.9	2	2	5 / 1	Birke	6.5	2	1	5 / 2
Poplar	45	6.9	1	2	6 / 2	Pappel	6.5	0	1	5 / 2
Willow	41	6.2	9	4	6 / 2	Weide	5.9	9	3	5 / 2
Holly	11	5.6	-	10	5 / 2	Ilex	6.2	0	1	4 / 2
Hazel	1	6.4	2	2	5 / 2	Haselstr.	6.2	1	0	12 / 3
Animate: four-footed animals										
Lion	225	8.0	17	8	4 / 2	Löwe	7.6	1	11	4 / 2
Pig	142	8.0	8	8	3 / 1	Schwein	7.8	10	5	7 / 1
Horse	348	7.9	117	105	5 / 1	Pferd	7.6	43	51	5 / 1
Deer	95	7.9	13	6	4 / 1	Reh	7.8	12	2	3 / 1
Rabbit	48	8.0	11	11	6 / 2	Hase	7.8	7	7	4 / 2
Ape	4	8.0	3	4	3 / 1	Affe	7.8	1	1	4 / 2
Fox	44	8.0	13	10	3 / 1	Fuchs	7.7	7	4	5 / 1
Bull	40	8.0	14	21	4 / 1	Ochse	7.7	78	1	5 / 2

continued overleaf

continued from previous page

English name	B&M	mean subjective animacy rating (E)	K&F	CELEX (E)	letters / syllables	German Name	mean subjective animacy rating (G)	Ruoff	CELEX (G)	letters / syllables
Animate: fish										
Trout	216	7.7	4	16	5 / 1	Forelle	7.8	4	1	7 / 3
Shark	176	8.0	1	14	5 / 1	Hai	7.8	0	1	3 / 1
Salmon	142	7.7	3	6	6 / 2	Lachs	7.7	0	0	5 / 1
Whale	78	7.9	1	6	5 / 1	Wal	7.8	0	0	3 / 1
Lobster	24	7.1	1	1	7 / 2	Hummer	6.9	0	0	6 / 2
Eel	17	7.9	2	4	3 / 1	Aal	7.8	0	5	3 / 1
Squid	8	7.8	-	1	5 / 1	Tintenfisch	7.8	0	0	11 / 3
Halibut	20	7.1	1	0	7 / 3	Heilbutt	7.2	0	0	8 / 2
Animate: birds										
Robin	377	7.7	2	12	5 / 2	Rotkehlchen	7.6	0	0	11 / 3
Parrot	72	8.0	1	3	6 / 2	Papagei	7.7	0	2	7 / 3
Hawk	111	7.9	14	4	4 / 1	Habicht	7.6	1	1	7 / 2
Sparrow	237	7.9	1	3	7 / 2	Spatz	7.8	2	1	5 / 1
Geese	12	7.6	3	2	5 / 1	Gänse	7.0	-	3	5 / 2
Falcon	28	7.7	4	1	6 / 2	Falke	7.5	1	2	5 / 2
Lark	15	7.5	2	3	4 / 1	Lerche	7.4	0	1	6 / 2
Raven	14	7.8	1	2	5 / 2	Rabe	7.7	0	1	4 / 2
Animate: insects										
Ant	258	7.7	6	4	3 / 1	Ameise	7.8	0	1	6 / 3
Beetle	161	7.8	1	5	6 / 2	Käfer	7.5	6	5	5 / 2
Spider	177	7.8	2	4	6 / 2	Spinne	7.5	0	5	6 / 2
Mosquito	227	7.6	1	2	8 / 3	Mücke	7.9	4	1	5 / 2
Caterpillar	30	6.5	-	2	11/ 4	Raupe	7.8	0	0	5 / 2
Hornet	33	7.4	1	0	6 / 2	Hornisse	7.7	1	0	8 / 3
Lice	12	7.4	2	2	4 / 1	Läuse	7.7	0	1	5 / 2
Cockroach	36	7.3	-	2	9 / 2	Kakerlake	7.3	1	0	9 / 4
Inanimate: clothing										
Shirt	352	1.4	27	45	5 / 1	Hemd	1.5	14	17	4 / 1
Socks	330	1.3	7	16	5 / 1	Strümpfe	1.4	3	0	8 / 2
Shoes	274	1.8	44	65	5 / 1	Schuhe	1.8	55	22	6 / 2
Dress	240	1.4	67	63	5 / 1	Kleid	1.8	40	53	5 / 1
Slippers	1	1.5	7	8	8 / 2	Hausschuhe	1.2	0	0	10 / 3
Boots	8	1.6	20	30	5 / 1	Stiefel	1.6	9	9	7 / 2
Apron	2	1.3	7	7	5 / 2	Schürze	1.2	2	3	7 / 2
Collar	1	1.5	17	18	6 / 2	Kragen	1.3	2	5	6 / 2
Inanimate: toys										
Doll	285	2.5	10	17	4 / 1	Puppe	2.2	4	5	5 / 2
Truck	125	2.0	57	25	5 / 1	Laster	2.0	13	2	6 / 2
Train	113	1.9	82	68	5 / 1	Zug	2.1	66	12	3 / 1
Blocks	98	1.1	37	24	6 / 1	Klötze	1.0	1	0	6 / 2
Football	13	1.9	36	32	7 / 2	Fußball	1.7	24	37	7 / 2
Swing	10	2.0	24	14	5 / 1	Schaukel	2.3	3	0	8 / 2
Cards	12	1.5	36	27	5 / 1	Karten	2.1	35	28	6 / 2
Ship	8	1.6	83	44	4 / 1	Schiff	1.9	27	52	6 / 1
Inanimate: kitchen utensils										
Fork	349	1.4	14	12	4 / 1	Gabel	1.3	17	4	5 / 2
Spoon	369	1.3	6	11	5 / 1	Löffel	1.2	3	6	6 / 2
Pan	242	1.4	16	22	3 / 1	Pfanne	1.2	2	2	6 / 2
Pot	205	1.3	28	23	3 / 1	Topf	1.2	1	7	4 / 1
Dish	42	1.3	16	21	4 / 1	Platte	1.2	9	9	6 / 2
Plate	34	1.3	22	37	5 / 1	Teller	1.2	3	9	6 / 2
Sifter	12	2.3	1	0	6 / 2	Sieb	1.4	1	1	4 / 1
Funnel	1	1.1	1	0	6 / 2	Trichter	1.2	1	1	8 / 2

continued overleaf

continued from previous page

English name	B&M	mean subjective animacy rating (E)	K&F	CELEX (E)	Letters / syllables	German name	mean subjective animacy rating (G)	Ruoff	CELEX (G)	letters / syllables
Inanimate: furniture										
Chair	440	1.4	66	103	5 / 1	Stuhl	1.7	9	26	5 / 1
Dresser	143	1.4	1	5	7 / 2	Anrichte	1.9	2	2	8 / 3
Table	408	1.4	198	244	5 / 2	Tisch	1.6	36	80	5 / 1
Desk	230	1.4	65	21	4 / 1	Schreibtisch	1.4	7	7	12 / 2
Bookcase	43	1.4	2	3	8 / 2	Bücherregal	1.7	6	0	11 / 4
Cabinet	41	1.5	-	43	7 / 3	Kommode	1.3	0	3	7 / 3
Picture	15	2.0	162	106	7 / 2	Bild	2.3	19	226	4 / 1
Mirror	11	1.5	27	41	6 / 2	Spiegel	1.5	5	36	7 / 2
Inanimate: carpenters' tools										
Saw	394	1.4	1	97	3 / 1	Säge	2.0	25	2	4 / 2
Screwdr.	214	1.5	-	15	10 / 3	Schraubenz.	1.6	2	6	15 / 4
Plane	147	1.8	114	45	5 / 1	Hobel	2.1	8	1	5 / 2
Chisel	103	1.6	4	2	6 / 2	Meißel	1.8	0	0	6 / 2
Pencil	37	1.3	34	15	6 / 2	Stift	1.7	0	2	5 / 1
Drill	52	1.3	81	2	4 / 1	Bohrer	2.1	1	0	6 / 2
Bolts	12	1.5	1	4	5 / 1	Bolzen	1.0	1	0	6 / 2
Wedge	25	1.5	4	7	5 / 1	Keil	1.7	21	3	4 / 1
Inanimate: stones										
Ruby	419	2.0	1	1	4 / 2	Rubin	1.8	0	0	5 / 2
Emerald	329	1.7	3	1	7 / 3	Smaragd	2.1	1	0	7 / 2
Sapphire	245	1.7	1	1	7 / 2	Safir	2.6	1	0	6 / 2
Garnet	44	1.9	1	1	6 / 2	Granat	2.0	0	0	6 / 2
Crystal	5	1.7	23	12	7 / 2	Kristall	2.1	1	1	8 / 2
Cameo	4	1.8	1	1	5 / 3	Kamee	1.9	-	-	5 / 3
Marble	2	1.6	21	22	6 / 2	Marmor	2.1	0	6	6 / 2
Moonst.	6	1.5	-	-	9 / 2	Mondstein	1.8	-	-	9 / 2
Inanimate: weapons										
Rifle	163	1.7	63	16	5 / 2	Gewehr	1.6	20	12	6 / 2
Knife	405	2.0	-	28	5 / 1	Messer	1.0	3	5	6 / 2
Sword	110	1.5	7	13	5 / 1	Schwert	1.6	2	7	7 / 1
Rope	84	1.6	-	14	4 / 1	Seil	2.0	1	3	4 / 1
Stone	20	1.2	58	41	5 / 1	Stein	1.2	68	32	5 / 1
Missile	28	2.0	48	26	7 / 2	Flugkörper	2.1	0	0	10 / 3
Arrow	20	1.7	14	8	6 / 2	Pfeil	2.1	0	7	5 / 1
Sabre	6	1.8	-	1	5 / 2	Säbel	1.6	0	1	5 / 2
Inanimate: musical instruments										
Drum	322	1.7	11	5	4 / 1	Trommel	2.1	2	5	7 / 2
Violin	271	1.8	11	4	6 / 3	Geige	1.7	1	1	5 / 2
Flute	246	1.9	-	4	5 / 1	Querflöte	2.1	1	2	9 / 3
Harp	105	2.0	1	2	4 / 1	Harfe	2.2	0	1	5 / 2
Recorder	5	2.2	7	0	8 / 3	Blockflöte	2.1	4	9	10 / 3
Bells	10	2.0	8	14	5 / 1	Glocken	1.9	4	0	7 / 2
Bugle	17	2.0	-	0	5 / 2	Waldhorn	1.6	1	0	8 / 2
Pipe	1	1.5	20	22	4 / 1	Pfeife	1.7	4	8	6 / 2

APPENDIX 11: Experiments 4a/4b, stimulus list.

Matching Parameters and Descriptives of Interlingual Homographs (IH) and Controls (C). (L = numbers of letters; S = numbers of syllables; KF = Kučera & Francis (1967) frequencies for English; Ruoff = Ruoff (1990) frequencies for German; CELEX E = CELEX frequencies for English; CELEX G = CELEX frequencies for German; Conc = concreteness). Phonetic similarity was calculated as per the method in Dijkstra et al., 1999.

Type	Item	L	S	KF	Ruoff	CELEX E	CELEX G	Conc	English transcription	German transcription	Phonetic similarity
IH	HOSE	4	1	9	26	3	11	596	həuz	ho:zə	0.5
C	DUCK	4	1	9		4		606			
IH	TAG	3	1	5	861	5	292	-	tæg	ta:k	0.3
C	JUG	3	1	5		8		-			
IH	SAGE	4	1	2	10	3	53	462	seidʒ	za:gə	0.0
C	WEED	4	1	1		8		600			
IH	BALD	4	1	5	30	8	161	-	bɔ:ld	balt	0.4
C	CUTE	4	1	5		3		-			
IH	HERD	4	1	22	17	9	7	565	hɜ:d	heɐt	0.3
C	PLUG	4	1	23		6		558			
IH	HERB	4	1	7	27	5	1	558	hɜ:b	heɐp	0.3
C	PEAR	4	1	6		2		634			
IH	ALTER	5	2	15	29	4	45	430	ɔltə	altɐ	0.3
C	UPSET	5	2	14		20		282			
IH	FORT	4	1	55	20	23	91	580	fɔ:t	fɔɐt	0.5
C	SEAT	4	1	54		78		568			
IH	MUSTER	6	2	3	15	1	20	-	mʌstə	mʊstɐ	0.5
C	DIGEST	6	2	3		1		-			
IH	GUT	3	1	1	1052	5	522	-	gʌt	gu:t	0.6
C	RIB	3	1	1		2		-			
IH	MUTTER	6	2	1	296	1	91	-	mʌtə	mʊtɐ	0.4
C	INFECT	6	2	1		1		-			
IH	NUN	3	1	2	250	5	824	583	nʌn	nu:n	0.6
C	TOY	3	1	2		15		567			
IH	BITTEN	6	2	3	3	5	54	-	bɪʔn	bɪtn	0.8
C	PRICED	6	2	4		1		-			
IH	STERN	5	1	23	3	9	34	349	stɜ:n	ʃtɛɐn	0.4
C	FAINT	5	1	25		25		462			
IH	DOSE	4	1	11	3	6	3	-	dəus	do:zə	0.3
C	MILL	4	1	11		10		-			
IH	ROMAN	5	2	58	2	38	48	-	ɹəumən	ʁoma:n	0.5
C	TRULY	5	2	57		38		-			
IH	HUT	3	1	13	22	22	14	589	hʌt	hu:t	0.6
C	FUR	3	1	13		19		601			
IH	RAT	3	1	6	7	9	75	624	ɹæt	ʁa:t	0.3
C	JAM	3	1	6		12		563			
IH	MIST	4	1	14	50	13	3	497	mɪst	mɪst	0.9
C	FORK	4	1	14		12		592			
IH	SPUR	4	1	13	2	4	28	-	spə:	ʃpu:ɐ	0.3
C	DICE	4	1	14		1		-			