Cognitive advantages in bilingualism:  
Is there a “bilingual paradox”?  

Antonella Sorace  
University of Edinburgh and University of Tromsø  

Abstract  
The term ‘bilingual paradox’ has been used to refer to the discrepancy between two opposite popular reactions: on the one hand, marvel at the ease with which young bilingual children acquire more than one language and, on the other hand, fear that they may experience confusion and delays (Petitto & Kolverman 2003). This paper deals with a different version of the ‘bilingual paradox’. Recent research has argued, somewhat counterintuitively, that the experience of growing up with two languages has positive effects on (some aspects) of non-linguistic general cognition (Bialystok 2009) but negative effects of (some aspects) of language competence. While most of the linguistic ‘disadvantages’ of bilingualism have been discussed in the domain of lexical competence, this question is approached here from the perspective of grammatical competence. The particular focus is on structures requiring coordination and rapid integration of syntactic and pragmatic information, such as of the use of pronouns and other anaphoric forms. It is suggested that the advantages and disadvantages of bilingualism can be regarded as two sides of the same coin, both in the linguistic and in the cognitive domain.  

1. Advantages of being bilingual  
One of the key findings from research on language and cognition in bilinguals is that knowing more than one language brings a wide range of advantages in childhood, in adulthood, and in old age. Children who know more than one language have a spontaneous understanding of language structure and therefore an enhanced ability to learn new languages. They also tend to have an earlier grasp of some essential background components of literacy, such as the invariance of print meaning and its symbolic function (Bialystok 2002). Moreover, bilingualism gives children advantages that go well beyond languages. These advantages are particularly evident in tasks that involve cognitive flexibility and the control of attention: bilinguals seem to be better at selectively paying attention, at inhibiting irrelevant information, and at switching between alternative solution to a problem ((Bialystok & Martin 2004;
Bialystok et al 2009). Importantly, these benefits do not appear across the board: bilinguals do not seem to have an advantage over monolinguals with respect to functions that depend on the way knowledge is represented, such as encoding problems or drawing logical inferences.

At the root of these cognitive advantages is the fact that bilingual speakers develop a powerful mechanism for keeping the two languages separate, so that fluency in one language can be achieved without intrusions from the language not in use. Therefore, the bilingual child’s constant experience of having two languages available and inhibiting one when the other is activated (Costa et al. 2008; Green 1998) enhances executive control in other domains. Recent studies show that some of these cognitive advantages are maintained in old age (Bialystok et al. 2004; Kavè et al. 2008) and that bilingualism may provide a ‘cognitive reserve’ against the decline of general processing functions both in normal and in pathological cognitive aging.

Are bilinguals more efficient at inhibition of irrelevant information, or do they have an enhanced ability to selectively activate relevant information? Do they acquire a range of more subtle advantages, such as the ability to ‘modulate’ executive function according to the type of task they engage in? While these possibilities are the focus of current research (see e.g. Blumenfeld & Marian 2011), there are two pieces of evidence that inhibitory control is indeed a crucial element of the bilingual advantage. The first comes from recent research on bimodal bilinguals, who know a signed language and a spoken language (Emmorey et al 2008). Bimodal bilingual do not need to inhibit one of the two languages because there is no competition for the same channel; in fact they tend to simultaneously produce signs and spoken words (code-blends) rather than code-switches (Emmorey et al., 2009). These bilinguals, unlike the speakers of two oral languages, do not show an advantage over monolinguals in executive control tasks. The second piece of evidence is provided by a study by Treccani, Argyri, Sorace & Della Sala (2009), who exploited the negative priming paradigm. Negative priming is a slower or inaccurate response to a target stimulus that had to be previously ignored: the stronger the inhibition necessary to ignore the distractor, the larger the negative priming effect (see Tipper 1985 on monolinguals). The prediction therefore was that if bilinguals are advantaged in terms of inhibitory control, then they would be more affected by negative priming than monolinguals,
and this is exactly what Treccani et al. found. Importantly, this study demonstrates that the effects of bilingualism may look advantageous or disadvantageous depending on the nature of the task.

2. Disadvantages of being bilingual

The advantages in cognitive control are accompanied by disadvantages in the domain of language (Bialystok 2009; Bialystok et al. 2009). The most obvious (and most studied) examples have been found at the level of lexical competence and access. Bilingual children may have some developmental advantages in word learning because they are less constrained by principles such a Mutual Exclusivity (Byers-Heinlein & Werker 2009) Although they often have smaller vocabularies in each language than their monolingual peers, initial vocabulary size differences are not detrimental to academic achievement (Barac & Bialystok 2011). However, since words of the unwanted language compete with those of the intended language, lexical selection for bilinguals is a more demanding process than for monolinguals. The lexicons of both languages are activated in bilingual production at all stages of the production process, even when speakers are in monolingual mode (Grosjean 2008). As a result, retrieval of words is slower in bilinguals of all ages than in monolinguals.

Are there disadvantages of bilingualism in other non-lexical linguistic domains? Let us explore the bilingual paradox with respect to a non-lexical phenomenon that has been robustly attested in different bilingual populations: the use of anaphoric expressions and, more specifically, the dependencies between pronominal forms and their antecedents.

3. Anaphoric expressions in bilinguals

Null subject pronouns in null-subject languages are syntactically licensed but their distribution is governed by discourse-pragmatic factors. The identification of pronouns is therefore dependent on the on-line computation of these factors in processing, as argued in recent syntactic theory (see e.g. Holmberg 2005; more on processing below). A robust finding that has emerged from research on adult second language (L2) acquisition is that native and non-native speakers of Italian diverge in their production and comprehension of pronominal subjects: this divergence is manifested in the greater optionality shown by non-native speakers (Belletti, Bennati
& Sorace 2007; Sorace 2003, 2005a,b; Sorace & Filiaci 2006a; Sorace 2011; see also Tsimpli & Sorace 2006 for data on L2 Greek). In production, near-native L2 Italian speakers optionally utter sentences such as (1b), with a ‘redundant’ overt pronoun, whereas a monolingual Italian speaker would produce (1c) with a null pronoun.

(1)  

a. Perchè Paolo è andato via così presto? ‘Why did Paolo leave so soon?’
b. Perchè lui non si sentiva bene
c. Perchè Ø non si sentiva bene
   “Because he wasn’t feeling well”

In contrast, errors involving null pronouns in inappropriate contexts are unattested; for example, pronouns are not omitted when a new referent is introduced (as in 2b), or when the sentence is explicitly contrastive (as in 3b).

(2)  

a. Perchè Paolo è andato via così presto? ‘Why did Paolo leave so soon?’
b. *Perchè Ø era noiosa  \(Ø = \text{la cena}\)
   “Because it (= the dinner) was boring”

(3)  

a. Maria ha detto che passava a prendere Paolo?  
   Maria has said that she was going to pick up Paolo?
b. *No, Ø ha detto che passava a prendere lei  \(Ø = \text{Paolo}\)
   No, Ø said that he was going to pick her up

The greater optionality of overt pronouns affects not only L2 speakers’ production, but also their interpretation of pronominal subjects. Their divergent interpretations are particularly clear with respect to intersentential anaphora involving a main clause that includes equally plausible antecedents and a subordinate clause containing an overt pronoun. In forward anaphora (where the main clause precedes the subordinate clause, as in (4)) Italian near-native speakers often interpret the overt pronominal subject of the embedded clause as coreferential with the lexical subject of the main clause (4a). In contrast, native speakers often interpret the overt pronoun in this context as referring to the complement (4b).

(4)  

a. Paola\(_{i}\) andrà a trovare Marta\(_{k}\) quando lei\(_{i}\) tornerà dalle vacanze.
b. Paola\(_{i}\) andrà a trovare Marta\(_{k}\) quando lei\(_{k}\) tornerà dalle vacanze.
Paola will go to visit Marta when she comes back from her holidays

In backward anaphora (i.e. with the subordinate-main clause order, as in (5)), native speakers typically interpret the overt subject as referring to an extralinguistic antecedent (5b); near-natives, on the other hand, have a strong preference for establishing a dependency between the overt pronoun and the matrix subject (5a).

(5)  
    a. Quando lei era in città, Paola è andata a trovare Maria.
    b. Quando lei era in città, Paola è andata a trovare Maria.

When she was in town, Paola went to visit Maria

A similar overextension of the scope of the overt subject pronoun is attested in other bilingual groups:

• First, it has been observed in native speakers of null subject languages in a situation of attrition due to prolonged exposure to a second language (see Tsimpili et al 2004 for Italian and Greek speakers who are long-term resident in an English-speaking country). The magnitude of the phenomenon, however, is smaller in L1 speakers experiencing attrition than in L2 speakers.

• Second, it has emerged from studies on German-English bilinguals (Wilson (2009) and Wilson, Keller & Sorace (2009). In German, both personal (er, sie, es) and demonstrative pronouns (der, die, das) can be used anaphorically. The division of labour between these anaphoric forms is similar to that between null and overt subject pronouns in null subject languages. As shown in (6), the pronoun identifies the subject of the matrix clause, whereas the demonstrative picks the complement:

(6)  
    Der Kellner erkennt den Detektiv als das Bier umgekippt wird. Er/Der ist offensichtlich sehr fleißig.

“The waiter recognizes the detective while the beer gets spilled. He-PRON/he-DEM is obviously hard working.”

Native German speakers, advanced L2 German speakers, and native German speakers who were long-term residents in the UK were tested on their comprehension of anaphoric forms using a visual world eye-tracking method. It was found that L2 learners’ preferences for the antecedents of pronouns are similar (though not
identical) to those of L1 speakers. In contrast, the dependencies of demonstratives are more indeterminate; L2 speakers have either no clear preference or a weak preference for a subject antecedent. Furthermore, demonstratives are significantly more affected by attrition than pronouns. The preferences of UK-resident German speakers for the antecedents of demonstratives show variability and divergence compared to monolingual Germans; either no clear preference, or a preference for a subject antecedent, depending on the length of residence in the UK. This pattern of convergence between L2 acquisition and L1 attrition in German-English bilinguals is thus the exact parallel to the pattern obtained for Italian subject pronouns in Italian-English bilinguals.

Third, developmental difficulty associated with the use of pronominal subjects is attested in bilingual L1 acquisition. Sorace, Serratrice, Filiaci, & Baldo (2009) conducted a large-scale study in which they compared two groups of school-age bilingual children acquiring two different combinations of languages; Italian-English (in which only one language allows null subjects) and Italian-Spanish (in which both languages allow null subjects). Elicited acceptability judgment experiments showed that both child bilingual groups accepted significantly more overt subjects referring to topic antecedents (as in Paperino; ha detto che lui; è caduto ‘Donald Duck; said that he; fell’) than monolingual children, regardless of language combination. Moreover, the younger monolinguals also did this significantly more often than the adult controls, indicating that these aspects of the syntax-pragmatics interface are acquired late (Sorace & Serratrice 2009).

The irrelevance of typological similarity with respect to the overuse of overt pronouns strongly suggests that language interference cannot be the only cause of this phenomenon. The pattern is also attested in adult bilingual speakers of two null subject languages of the same type: examples (7) and (8) are from a study of low and intermediate proficiency Spanish learners of Italian by Bini (1993):

(7) a. Quanti anni ha Pedro?  
How old is Pedro?  
b. Lui ha ventitre anni  
He is twenty-three

(8) Mia sorella e mio cognato escono per il lavoro e loro lavorano a Paseo de la
Other more recent studies of L2 speakers of two null subject languages, even if limited in number, show that overt pronouns are overextended particularly when the null-subject L2 is a second language (Malgaza & Bel 2006; Lozano 2006; Mendes & Iribarren 2007), but not when it is a third language (Kraš 2008; Montrul et al. 2008). There is thus a potential discrepancy between L2 and L3 acquisition data, or between bilinguals and multilinguals – a point to which we will return.

4. Possible explanations (1): bilingual language representations and processing

Theoretical linguistic accounts often assume a distinction between syntax in a narrow sense and interfaces between syntax and other (semantic, discourse, lexical) cognitive systems (Avrutin 1999, 2004; Burkhardt 2005; Ramchand & Reiss 2003). According to this view, features in ‘narrow syntax’ are responsible for parametric differences among languages; features at interfaces (such as [Topic Shift] and [Focus]) have interpretive effects, as they can be ‘read’ by the conceptual/intentional systems of cognition. These two types of features, within this view, are dissociated in bilingual speakers: syntactic features are acquirable and are unaffected by attrition, but interface features are not. One possible reason is that these features remain, or become, underspecified in the bilingual competence, and this underspecification generates optionality. The optionality in bilingual speakers therefore involves interpretable features linked to a parametric choice that differs between the L1 and the L2. An interpretable feature that is specified in the null subject language (regardless of whether this is L1 or L2) for a particular syntactic structure remains, or becomes, unspecified due to the absence of a similar feature in the other language.

This explanation, however, is unable to account for the fact that overt subject pronouns are overextended by bilingual speakers of two similar languages. Let us thus approach the problem from the processing perspective and suppose that optionality is related not (only) to the underspecification of the interface features regulating the mapping between levels of representation, but rather to the integration of information from different domains, which must take place in real time whenever a pronoun (or
other anaphoric form) is produced or encountered in a particular context. The efficiency of integration may remain, at least occasionally, problematic for bilingual speakers. In this perspective, structures that require not only knowledge of different types of information but also the ability to coordinate them in real time are more ‘complex’, or more costly in processing terms, than structures that require only one type of information. It follows that processing interface structures should in general be more costly than processing narrow syntax structures. Referential subject pronouns in Italian, in fact, qualify as ‘complex’, since they demand mastery of both morphosyntactic properties and discourse conditions. Furthermore, even monolingual speakers, or bilingual speakers of languages that do not differ in terms of syntax-discourse interface mappings, should occasionally experience integration problems.

Indeed, recent psycholinguistic research on anaphora resolution in native speakers of null-subject languages lends support to this argument. Carminati (2002, 2005) provides experimental evidence that null and overt pronouns in Italian have distinct and complementary functions, manifested in their distinct biases for antecedents in different syntactic positions. Null pronouns have a strong bias towards an antecedent in Spec IP (normally – but not exclusively – the subject), whereas overt pronouns prefer an antecedent in positions lower in the phrase structure (normally – but not exclusively – a complement): this is referred to as the ‘Position of Antecedent Strategy’ (henceforth PAS). Thus, the initial antecedent assignment in anaphora resolution is structurally based, but if the predicate is pragmatically biased to contradict the PAS, reanalysis is necessary at a measurable processing cost. For example, the sentence in (9b) elicits longer reading times than the sentence in (9a).

Similarly, phi-features (gender, person and number) may be consistent or inconsistent with the initial parse based on the PAS: incongruences between the PAS and these features (as in (10a) with an overt pronoun and (10b) with a null pronoun) are costly in processing terms (examples 9-11 are adapted from Carminati 2000).

(9)   a. Quando Paola, ha telefonato a Luisa, ø, le ha dato la buona notizia.
    ‘When Paola telephoned Luisa, she gave her the good news.’
   b. Quando Paola aiutò Luisa, ø, era svenuta.
   When Paola aided Luisa, she was unconscious.
The PAS is a highly efficient processing principle that belongs to the interface between syntax and discourse. As Carminati herself suggests, not only is there a reliable correspondence between the structural position Spec IP and the notion of topic, but also pragmatic principles are the core of antecedent preferences. So, for example, using an overt pronoun to refer to a topic antecedent would represent a violation of Grice’s maxim of quantity, because since another form – the null pronoun - is available for the same purpose, the comprehender assumes that it should have been used instead. Crucially, however, there is a difference between null and overt pronouns with respect to the strength of the PAS. Carminati’s experimental data indicate that while the preference of the null pronoun for subject antecedents is very robust, the overt pronoun shows more flexibility in its antecedent preferences: a weaker processing cost may be incurred if an overt pronoun takes a subject antecedent than if a null pronoun takes a non-subject antecedent. The antecedent preferences of overt pronouns appear to be sensitive to contextual factors: the monolingual grammar is more tolerant of PAS violations in unambiguous sentences, in which the potential for miscommunication is low. It appears, therefore, that monolingual speakers may be occasionally unable or unwilling to engage in full processing when they know that the context is sufficiently unambiguous, as in (11b), in which there is only one referent, or (11c), in which the pronoun agrees in number with only one of the two antecedents; in these cases they may produce a sentence with an unnecessary, or redundant overt pronoun which does not impair antecedent assignment in comprehension. An overt pronoun would be much less likely to be produced in the ambiguous context of (11a).

(11) a. Paola, passava molto tempo con Luisa quando lei era in vacanza.
‘Paola, used to spend a lot of time with Luisa when she was on holiday.

b. Giorgio, ha detto che lui non voterà alle prossime elezioni.
‘Giorgio, has said that he will not vote at the next election’.

c. Quando Carlo, ha visto i suoi amici, lui era molto contento.
‘When Carlo saw his friends he was very happy’

Thus, overt pronouns may be used inappropriately when the speaker does not pay enough attention to encoding the utterance from the comprehender’s perspective, or is otherwise unable to do so when, for example, the processor is overloaded: in this case, the PAS is relaxed, although comprehensibility is not compromised. It is plausible to think that bilingual speakers, whose processing resources are more restricted, may resort to relaxing the PAS in a wider range of contexts and with less consideration for overall potential ambiguity. The overt pronoun may therefore be a kind of default form used to relieve processing demands when these become temporarily unmanageable. If these assumptions are correct, one would expect that these patterns of pronoun overgeneralization in Italian should be produced not only by non-native speakers whose L1 is English, but also by non-native speakers from different language background, including languages that have a similar pronominal system to that of Italian: exactly what emerges from the studies just reviewed.

5. Possible explanations (2): locating the bilingual processing problem

But what exactly is the bilingual processing problem, and why is it of significantly greater magnitude than in monolinguals? Two hypotheses will be outlined: the first relies on competition for processing resources; the second on inconsistent ability to integrate pragmatic/contextual information. Pronominal use and other syntax-pragmatics interface structures require not only the ability to choose the correct pronoun-antecedent mapping according to the current assessment of the interlocutor’s perspective and the referent’s conceptual status (inhibiting the inappropriate ones both within and across languages), but also the ability to update the discourse model dynamically following a language switch or a change in the referent’s status (for example, whether it has been recently mentioned and is therefore still prominent for both speaker and interlocutor.

The use of anaphoric expressions is therefore consuming in terms of processing resources and sensitive to cognitive load. Discoordination in pronominal reference has emerged as a phenomenon not only in bilinguals but also in other populations sensitive to cognitive load, such as ageing speakers (Titone et al 2000), schizophrenic patients (Phillips & Silverstein 2003), and autistic children (Arnold, Bennetto & Diehl
The interesting result in Arnold, Bennetto & Diehl’s study is that high-functioning autistic children tend to overspecify anaphoric references (e.g. they use more explicit expressions), just like bilinguals.

As discussed above, bilinguals need to exercise executive control to avoid interference from the unwanted language. Suppose that anaphoric dependencies partly draw on the same pool of attentional resources used to keep the two languages separate: this creates a competition for resources when bilinguals engage in linguistic tasks that are sensitive to cognitive load, which may impact on different aspects of the task. In the case of anaphoric dependencies, the assessment of the interlocutor’s knowledge state and of the relative accessibility of referent may be (inconsistently) problematic: adult speakers do not reliably consider what the interlocutor knows in their initial encoding of referential expressions (Keysar, Lin & Barr 2003) and resources are needed to recover from initially ‘egocentric’ computations. Asymmetric inhibition effects (Meuter & Allport 1999) may account for the size of the phenomenon of overextension of overt pronouns: recall in fact that this is smaller in L1 Italian speakers under attrition than in L2 speakers of Italian. In L2 speakers, the unwanted language is their (still dominant) L1, which requires more resources to be inhibited. In L1 speakers experiencing attrition, in contrast, the unwanted language is their (less dominant) L2, which requires fewer resources to be inhibited. A third and so far unexplored possibility is that increased inhibitory control and less efficient integration ability may be two sides of the same coin, in a similar way to the relationship between inhibitory control and negative priming. The necessity to integrate pragmatic information and to update the current mental representation of the anaphoric context may be regarded, in a sense, as ‘the opposite’ to the ability to selectively focus attention and inhibit irrelevant details: the enhancement of the latter may be at the expense of the former. The difference is likely to be affected by age of onset of bilingualism and/or balance between the two languages. It has been argued (Costa & Santesteban 2004) that there may be differences between early and late bilinguals (or between more balanced and less balanced bilinguals) with respect to the presence or the type of effects of the bilingual experience on executive function: one of the differences may be not so much the absence of an advantage but its restriction. Preliminary results (Bak, Everington, Garvin & Sorace 2008; Bak & Sorace, submitted) support the hypothesis that late bilinguals may have an advantage in
inhibitory control but not in task-switching and adapting to new conditions. These researchers employed three tests from the Test of Everyday Attention (TEA, Robertson et al, 1994) of increasing complexity, which measured (from least to most complex) sustained attention, selective attention and inhibition, and task switching and monitoring. The battery was administered to monolinguals, early consecutive bilinguals, and late bilinguals. While there were no significant differences between early and late bilinguals, the advantage for late bilinguals was significantly larger than monolinguals for the inhibition test but not for the switching and monitoring test. One possible account is that early bilinguals acquire the ability not only to apply inhibitory control, but also to ‘disengage’ inhibition when required by the nature of the task; disengagement of inhibition allows more flexibility in task switching and facilitates updating of the mental representation of the problem. Further research on bilingual and multilingual speakers of different language combination is necessary to establish whether disengagement of inhibition might be at work also for the use of anaphoric expressions and whether it may be in part responsible for the different extent to which bilinguals and multilinguals resort to the use of overt pronouns as a default.

6. Conclusions
The experience of handling two languages has a significant impact on linguistic and general cognitive abilities. In both domains, however, this impact may translate into advantages or disadvantages according to the nature of the situation and the task faced by bilinguals. In several cases, the effects of bilingualism may magnify processing difficulties that are also experienced by monolinguals. Examining linguistic behaviour from the point of view of the interaction of language and general cognition reveals that problems traditionally regarded as exclusively linguistic may be at least in part due to the pressure of controlling two languages in real time. Much more research is necessary to understand these issues, but for the time being the ‘bilingual paradox’ is perhaps less paradoxical than it appears at first sight.
References


Margaza, P. & Bel, A. 2006. Null subjects at the syntax-pragmatics interface:


