Determining language contact effects in ancient contact situations

Sarah Thomason
University of Michigan

Proving the existence of ancient language contacts is easy; proving the existence of ancient contact-induced language change is much more difficult, by comparison to analyses of modern contact situations. This paper surveys some ancient contacts and their effects on the languages. The main conclusion is that the historical methods used for analyzing better-documented contact situations can be applied to ancient contact situations as well. But the chances for success are likely to be more limited, because gaps in the available information may make it impossible to satisfy the prerequisites for proposing contact-induced changes.

1. INTRODUCTION. Language contact is almost as old as humankind. It has surely been a constant feature of human culture for as long as humans have spoken more than one language—which, if we assume a monogenetic origin for humans and therefore for human language, would presumably have been somewhere from several hundred to several thousand years after the beginning, depending on when the earliest speech community broke up into two or more subcommunities or separate communities.

Evidence of language contact, however, is much more recent. To a certain extent we can infer the existence of language contacts from known population movements and cultural practices. So, for instance, the amount of linguistic diversity in New Guinea, with its thousand or so languages, must have taken many millennia to develop (even before Austronesian speakers arrived a few thousand years ago to settle around the island’s coast), and there is no reason to suppose that the intensive language contact that characterizes the island is a modern phenomenon.

But until the invention of writing, or at least until the earliest writing to survive to modern times—namely, Sumerian writing from the late fourth millennium BCE—we have only vague evidence of contacts among peoples and their languages. Some of the early written
evidence is indirect; in the second millennium BCE, for example, the Egyptians had military forces and administrators outside Egypt itself, and they had foreign prisoners and forced laborers inside Egypt (Kammerzell 1998:97). This means that speakers of other languages must have interacted with speakers of Egyptian.

Most of the documentary evidence is direct, and it shows clear evidence of contacts between languages in all cultures touched by literacy, including cultures with their own writing and cultures that were written about in ancient times. Consider the famous Gilgamesh Cycle, a Babylonian epic poem from the second millennium BCE: it originated as a Sumerian epic, spread to Babylonia and was translated and adapted into the Semitic language Akkadian there, and then spread to the Hurrians, the Hittites (who spoke the oldest attested Indo-European language), and other ancient Near Eastern peoples (Kramer 1981). Its main importance is literary and cultural, but its spread obviously involved extensive language contact.

By ca. 2500 BCE, the northern part of Babylonia is believed to have had a bilingual Sumerian/Akkadian population, while the southern part was still mostly Sumerian-speaking. Akkadian was the official language of the Akkadian dynasty of Sargon (2234-2154 BCE), but many inscriptions were bilingual; later, but still in the late third millennium BCE, many official documents were written in Sumerian, even as Sumerian declined as a spoken language (Böck 2004). Akkadian, meanwhile, became the major lingua franca of the ancient Near East (Márquez Rowe 2004), from the Hittite empire—where scribes wrote in Akkadian because it was the language of diplomacy (Gary Beckman, personal communication, 2003)—to Egypt, where diplomatic correspondence in Akkadian was found (Kammerzell 1998:97).

Later, but still in the ancient period, direct evidence of language contacts proliferated. Bilingual inscriptions were found in various places. The most famous of these is the Rosetta Stone, from 196 BCE, which has a single text in Egyptian hieroglyphics, Demotic Egyptian, and Greek; this was the key to the decipherment of the Egyptian hieroglyphs. In his autobiography, Darius the Great (550-486 BCE, reigned 522-486)—or rather his ghostwriter—wrote that, in addition to cuneiform writings in Persian, Elamite, and Akkadian, ‘I made inscriptions in other ways, in Aryan, which was not done before’ (Olmstead 1948:116; ‘Aryan’ is another name for the Indic subbranch of the Indo-European language family). In the 5th
century BCE, Herodotus (ca. 485-425 BCE) wrote what may be the very first report of contact-induced structural change (Book IV, ch. 117): ‘The Sauromatae speak the language of Scythia, but have never talked it correctly, because the Amazons learned it imperfectly at the first’—a reference to the speech community that arose from unions of Amazons with Scythian men.

A possible rival for the earliest report of contact-induced change, also involving mixed marriages, is in the Old Testament of the Bible, in Nehemiah 13:23-24: ‘In those days also saw I Jews that had married wives of Ashdod, of Ammon, and of Moab: And their children spake half in the speech of Ashdod, and could not speak in the Jews’ language, but according to the language of each people.’ Nehemiah served either King Artaxerxes I of Persia (465-424 BCE) or King Artaxerxes II (404-359 BCE), but further complications make this Biblical passage hard to date with confidence. During the first half of the third century BCE the Pentateuch (the first five books of the Old Testament) was first translated into Greek, in Ptolemaic Alexandria, obviously by bilingual translators (Fernández Marcos 2004).

Outside the ancient Near East and Mediterranean, we find more recent but still quite old examples of language contact. Speakers of the Chamic languages of Southeast Asia, for instance, returned to mainland Asia about 2,000 years ago and were in contact with speakers of unrelated languages before that time; contact is documented from the earliest period of attestation, which began with inscriptions from the fourth century CE (Thurgood 1999). Japanese underwent very extensive cultural and linguistic influence from Chinese, borrowing a great many words (including an entire set of numerals), especially during the sixth century CE. Arabic incorporated influences from other languages as a result of its spread, starting in the late seventh century CE (Versteegh 1984; see also al-Bakri and Ibn Khaldun). In the New World, Quechua, the official language of the Inca empire, was in contact with other Andean languages before the Spanish invasion in 1532 (Mannheim 1991). The Qing dynasty in China (1644-1911 CE) was Manchu-speaking, but outside the imperial court people spoke Chinese languages. Many more examples of old language contacts can be found all over the world; the ones listed here are a tiny sample.

2. Ancient Loanwords. Loanwords are often, though not always, easy to detect. For one thing, they tend to cluster outside the basic vocabulary. Even in intense contact
situations in which a language borrows a great many words, including basic vocabulary items, most loanwords will still be in the non-basic vocabulary. In a two-language contact situation, if at least one of the languages is well understood, identifying loanwords will probably be a straightforward process. For example, there are a few Egyptian loanwords in Ancient Greek literature, e.g. in Herodotus and Aeschylus; but these reflect sporadic contacts for purposes of trade and the like, probably not widespread bilingualism, and the loanwords designate specifically Egyptian items (Torallas Tovar 2003). By contrast, lexical borrowing in the other direction, from Greek into Egyptian, a few hundred years later, was massive: Reintges (2004) estimates that perhaps 40% of the Coptic lexicon was borrowed from Greek, including many function words. Older Egyptian also has many loanwords, but from other languages, especially Northwest Semitic (e.g. words for ‘shield’, ‘apple’, and ‘wool’), from the mid-second millennium BCE; and it has a few loanwords from other languages by the end of that millennium—from Akkadian (‘sofa, bed’; ‘cart’), Sumerian (‘temple, palace’), and Hurrian (‘warrior’; ‘cart-driver’) (Kammerzell 1998).

In Anatolia, Hittite adopted numerous loanwords from neighboring languages. (Few loanwords seem to have traveled in the opposite direction, although Melchert (2000[1995]:2152) reports that there are a handful of Hittite loanwords in Assyrian texts from the nineteenth century BCE, among them a word for ‘obligation, contract’.) Loanwords in Hittite are generally confined to ‘terms relating to aspects of the cult, items of higher culture, and the names for some flora and fauna’....At least 75% of the core vocabulary is based on inherited Indo-European material’ (ibid.). Many of the loanwords are from Hattic (Beckman 1996:24), among them words for a kind of bread, an epithet of the sun-god, ‘throne’, and ‘woe, pain’ (Melchert 1994:35, 171, 195). Other loanwords in Hittite are from Akkadian (e.g. ‘tablet’, ‘exorcist’—Melchert 1994:21, 61, 170, 225), Hurrian (e.g. ‘door’, ‘slanderous’—ibid., pp. 170, 171), and Luwian, which was closely related to Hittite.

In all these cases, both the receiving languages and the source languages are well enough known to make the identification of loanwords non-problematic. It sometimes happens, however, that a loanword can be identified even when the source language is completely unknown. A famous example is the common Indo-European word for ‘hemp’, which is discussed by Sihler (1995:146). This word appears in Greek as kánnabis, in Old English as
hænep (the direct source of Modern English hemp), and in Middle High German as hanef. Latin cannabis is a loanword from Greek. English, where it means something rather different from ‘hemp’, later borrowed cannabis from Latin. The word must be reconstructed for Proto-(Western)Indo-European, because it is clearly quite ancient in the family, but for two reasons it must be an ancient borrowing rather than a native word: first, with three syllables it is too long for a native Indo-European root, and it is unanalyzable as a morphologically complex Indo-European word; and second, the plant was ‘probably not native to the PIE world’ (Sihler, ibid.). But no one (as far as I know) has discovered a plausible source for this ancient Indo-European loanword.

In the ‘cannabis’ case, although the donor language is unknown, the receiving language—Proto-Indo-European—has been studied so long and so successfully that we can be confident that the word is not native. Serious problems can arise in efforts to detect ancient borrowings, however, if neither the receiving language nor the proposed donor language is well understood. As Thurgood observes in his excellent case study of ancient and more recent borrowings in Chamic, an Austronesian group, ‘Without a reconstruction, it is often not possible to differentiate between similarities that are genetically-inherited and those that are contact-induced’ (1999:4). Thurgood discusses partial solutions in a particularly apt passage (1999:308-309):

‘Establishing that a MK [Mon-Khmer] form reconstructs to PC [Proto-Chamic] is done almost exclusively on correspondence patterns within PC...A number of specific vowels, certain diphthongs, particular consonants (e.g. the implosives, with a small number of well-known exceptions), and otherwise unique clusters only occur in MK borrowings. The fact that a word contains one of these elements is by itself evidence that it is a MK borrowing; however, it is almost never necessary to depend upon this evidence alone, as when a word contains one of these elements, it is inevitably the case that the word also lacks an Austronesian etymology.’

Part of the problem with identifying old Mon-Khmer loanwords into Chamic is that historical investigations of the relevant Mon-Khmer languages are not plentiful. Austronesian
has been intensively studied historically, but the Chamic group has not been, until Thur-good’s work. Firm conclusions therefore cannot be drawn about the origins of many Chamic words that lack Austronesian etymologies and also lack the tell-tale non-native sounds and sound patterns that point to a Mon-Khmer source. This problem of course also arises in many recent and even current contact situations; but it is likely to be more acute for ancient contacts, where additional data is much less likely to become available. Even so, as we will see in the next section, efforts to establish lexical interference are more often successful than efforts to establish structural interference.

3. ANCIENT STRUCTURAL INTERFERENCE. If two languages in contact share some structural features, this fact does not in itself provide evidence that the shared structures were transferred from one language to the other: the features may be shared by accident, where “accident” is construed broadly enough to include the workings of universal structural tendencies, i.e. markedness. Even if a language has borrowed many words from another language, it isn’t safe to assume that it has also borrowed structural features from that language. English, in spite of its thousands of French loanwords, has undergone at most moderate
structural interference from French, for example (Thomason & Kaufman 1988:263ff.).

Conversely, even if a language has no loanwords at all from a language with which its speakers were once in close contact, it isn’t safe to assume that there has also been no structural interference. The reason is connected with a robust sociolinguistic predictor for kinds of interference—namely, the extent to which the agents of contact-induced change are fluent in the receiving language. (The following discussion of this distinction, and of the method for proving that contact-induced change has occurred, is based on Thomason 2001, chapter 4, where examples and evidence can be found; the original source of the distinction, though in a slightly different form, is Thomason & Kaufman 1988.) In cases where the agents of change are fully fluent in the receiving language, imperfect learning plays no role in the transfer process. Let us call this type of contact-induced change **borrowing**. Typically, in borrowing, a speaker introduces features from his/her second language (L2) into his/her first language (L1), though the reverse direction of transfer also occurs. In borrowing situations, if there is any interference at all (which depends on such social factors as the intensity of contact and speakers’ attitudes), non-basic vocabulary will normally be the first type of linguistic feature to be transferred. If contact is intense enough, structural features and basic vocabulary may also be borrowed.

In sharp contrast, the predicted interference features are reversed if imperfect learning does play a role in the process: phonological and syntactic features predominate, and there may be little or no lexical transfer. (Inflectional morphology is least likely to be transferred in both types of interference.) When imperfect learning enters the picture, I call the process **shift-induced interference**, though sometimes there is no actual shift of one population to another group’s language because the L2 learners maintain their original L1 for in-group usage. At least three kinds of things happen in shift-induced interference. First, members of the shifting group may carry over features of their L1 into their L2, the target language (TL). Second, they may fail to learn certain TL features, especially features that are highly marked and thus harder to learn. These two phenomena combine to form TL\(_2\), the learning group’s version of the TL. Third, if the shifting group and the original speakers of the TL coalesce into a single speech community, the original speakers may adopt a subset of TL\(_2\) features, so that the merged community will eventually be speaking a third variety, TL\(_3\).
In most shift situations there will be at least a modest amount of lexical transfer from the shifting group’s L1 into the TL, their L2. This is especially likely if the original TL community has invaded the shifting group’s territory, because the newcomers will need words for unfamiliar flora and fauna, and probably also for some cultural items that are new to them. In a current or recent contact situation, some of those loanwords will surely still be found in the TL, along with the kinds of structural interference features that are typical of shift situations. But if the period of contact was in the distant past, it may well be that few or none of the loanwords will have survived in the modern language(s). Because many linguists assume that loanwords are the first and primary interference features in all kinds of contact situations, they tend to be reluctant to accept hypotheses of structural interference in the absence of numerous loanwords. This is (in my opinion) a major reason for the continuing controversy over whether ancient Indic acquired some structural features from Dravidian, e.g. retroflex consonant phonemes and agglutination in noun morphology (Emeneau 1956), when Dravidian speakers in the North of India shifted to the language of the invading Indic speakers. If ancient Dravidian loanwords were more plentiful in Indic, skeptical scholars might be more willing to accept the possibility of structural interference as well (see Thomason & Kaufman 1988:139-144 for more detailed discussion of this case).

Moreover, in some ancient contact situations there are no modern direct descendants of the old TL, the receiving language, so that we are completely dependent on ancient documents for our knowledge of the language. Sumerian, Hittite, Hurrian, Hattic, and Akkadian come to mind, but these are not the only examples. In such a case, we are very lucky indeed if a few loanwords left over from a shifting group’s original L1 happen to be attested in the surviving documentation for the old TL. It’s true that loanwords are often very numerous in cases of superstrate shift, as in the shift of the French-speaking Norman ruling class to English in England; but most shifting groups are not socioeconomically dominant over the TL community.

Because shift-induced interference may not leave evidence of contact in the form of easy-to-spot loanwords, the presence of loanwords cannot be a prerequisite for the existence of contact-induced structural change. And there is another reason for rejecting loanwords as a necessary precursor to structural interference, even in cases where the potential agents
of change are fully fluent bilinguals: in some cultures, lexical borrowing is frowned on, and it does not occur. But structural borrowing can occur in such contexts, according to Aikhenvald’s analyses of the Arawakan language Tariana of Brazil (1996, 2003). This makes the prediction about interference features in borrowing situations conditional on attitudinal factors that can block lexical borrowing, though the prediction holds in the vast majority of cases.

What, then, are valid and necessary requisites for establishing a claim of contact-induced structural change? First, one must identify a source language, or—in a case of shift in which all source-language speakers shifted to the TL—a language family to which the proposed source language belonged. In order to identify a source language, the analyst must make a case for contact intense enough to lead to interference. This is usually easy in borrowing situations, because loanwords are a clear sign of contact. It may be harder in shift situations, especially if the source language has vanished.

Second, one must identify shared structural features in the proposed source language, A, and the proposed receiving language, B. These features need not be (and often aren’t) identical in the two languages, but they must be similar enough to make a case for reinterpretation in B. The shared structural features should range over a variety of grammatical subsystems, to avoid the pitfall of interconnected features, e.g. aspects of morpheme ordering (Greenberg 1963).

Third, one must prove that the shared features are innovations in B—that is, that B has changed. Fourth, one must prove that the shared features are old in language A—that is, that A has not changed with respect to these shared features. Fifth and finally, it’s important to search for internal motivations for the innovations in B, because multiple causation is common enough that that possibility should never be overlooked. In other words, establishing that a given set of changes is motivated by contact does not mean that all the changes were triggered by contact alone, in other words.

If the first four requisites cannot be satisfied, then the case for contact-induced change cannot be considered firmly established. Indeterminacy of results is of course a fact of life for all historical linguists, so this is nothing new. But the likelihood of failure is greater with ancient contacts than with modern ones, again because there is usually much more
data available for modern contact situations. If ancient contact-induced changes were shift-induced, much of the evidence, both structural and lexical, is liable to have vanished over time and/or to be unattested in the documents that provide the only direct attestations of a language like Sumerian or Hittite. Making a case for contact-induced change in a language like Sumerian, which has no known ancient or modern relatives, will be the most difficult; a language like Hittite, which belongs to a branch of Indo-European that died out more than 2,000 years ago, at least has distant relatives and is therefore fairly well understood historically.

Cases for contact-induced structural change that are weakened by the failure to satisfy one or more of the requisites may still be quite promising, and they may offer the best historical explanation for the facts, even when they are incomplete.

The easiest kinds of structural interference to find and establish are phonological innovations that enter the receiving language with loanwords. In casual contact situations, the typical treatment of all loanwords is nativization—adaptation of the borrowed words to the receiving language’s native phonology. But in more intense contact situations, some foreign phonological features may be retained in loanwords, even if they never penetrate into the native vocabulary. Thurgood’s Chamic data, as we saw above, contains a number of foreign segments in borrowed words. Another typical example has been found in Palaic, a long-extinct Anatolian language of the Ancient Near East that was closely related to Hittite. According to Melchert, Palaic has loanwords from Hattic, and in these loanwords one finds /f/, which is foreign to Palaic (2000 [1995]:2154). This example satisfies all but one of the four crucial requisites: the proposed source language is known; the existence of contact with the receiving language is demonstrated by the presence of Hattic loanwords in Palaic; the shared feature in question is the presence of /f/ in both languages; and both Proto-Indo-European and the Anatolian branch of the family lack a phoneme /f/, so /f/ is definitely an innovation in Palaic. But the history of Hattic is unknown, and it also has no known relatives, so that we can’t absolutely prove that /f/ is old in Hattic. Nevertheless, given the fact that Palaic has /f/ only in Hattic loanwords, the case for borrowing must be considered strong even in the absence of evidence about the age of /f/ in Hattic.

A more complicated (and therefore more interesting) example in this same general cat-
egory is discussed by Kammerzell (1998:119ff.). In a section entitled ‘Kontaktinduzierte Suspension phonologischer Regeln?’ (‘Contact-induced suspension of phonological rules?’), Kammerzell describes several phonotactic rules in older Egyptian that barred certain consonant combinations from co-occurring within a single morpheme. Two obstruents of the same position class never co-occurred, for instance, and a voiceless non-emphatic stop never co-occurred with a voiced stop. But in later Egyptian these rules are violated in loanwords, e.g. brq ‘sparkle’, a loanword from Semitic. (Egyptian writing does not represent vowels, so the vowels are unknown.) As in the Palaic example, the structural borrowing is confined to loanwords. Unlike the Palaic example, however, this one fulfills all the requisites for establishing a change as contact-induced: the first three are parallel to the Palaic case, the innovation of the consonant combinations in Egyptian is attested in real-time documentation, and Semitic is well understood historically.

These are not, of course, the most dramatic kinds of contact-induced change. Convincing case studies of ancient innovations that are not confined to loanwords—especially syntactic changes—can be found in the literature. Kaufman’s 1974 study of mutual influences between Akkadian and Aramaic is one example; another is Watt’s investigation of Semitisms in the Greek New Testament (1997; see also Evans 2001). The widespread use of Akkadian and then Aramaic as lingua francas in the ancient Near East makes these two languages prime loci for contact-induced changes, both as donors and as recipients. As Márquez Rowe observes (2004), ‘Since Akkadian was not the native language in [Hattuša, Cyprus, Egypt, and elsewhere], this lingua franca was influenced by the spoken language of the scribes’—namely, via shift-induced interference, even if the scribes did not actually shift to Akkadian. And Khan (2004), discussing contacts between Aramaic and Iranian, notes that ‘[a]lready in the 5th century BC there is evidence of extensive influence of Iranian on eastern Aramaic. The convergence between Iranian and Aramaic is particularly evident in the modern spoken Aramaic dialects in the region of Kurdistan.’ In all these cases the requisites for establishing contact-induced change can be met, in spite of the long-past period of contact, because all the languages involved have been extensively analyzed historically.

A example that is comparable to Márquez Rowe’s is Vierros’s analysis of relative clauses in the Greek written by Hermias, an Egyptian notary in Pathyris from 109 to 98 BCE.
According to Vierros (2003:18), all of Hermias’s Greek texts show considerable influence from Coptic, which was, she argues, his native language. In relative clauses, the relative pronoun and the verb tend to agree in number and (in the pronoun only) gender with the subject of the relative clause rather than with the antecedent, as would be expected in native Greek. Vierros proposes that the source of this pattern lies in Demotic Egyptian, in which the most common relative clause formation consists of a ‘relative converter’ \( nt \) combined with a morpheme \( iw \) ‘circumstantial form’ and then with a suffixed pronoun that ‘indicates, or actually is, the subject of the relative clause’. This construction was apparently equated by Hermias with the Greek relative pronoun and inflected accordingly—correctly for Coptic, but incorrectly for Greek. The Greek/Egyptian contact situation was in fact one of mutual influence. Extensive borrowing interference in the other direction, from Greek to Coptic, has been described by Reintges (2004), including much lexical borrowing together with grammatical borrowing, discourse organization, and what Reintges calls the ‘Hellenisation of Egyptian syntax’. In these cases, as in others described in this section, the histories of the languages are so well known that the facts of interference can be established with confidence.

It should be noted that the non-native features described by Márquez Rowe and Vierros in the writings of scribes in Akkadian and Greek, respectively, cannot be assumed to be consistent, stable interference features in the languages as written (and perhaps spoken, at least in the case of Greek) by whole communities. But for purposes of understanding processes of contact-induced language change, it isn’t crucial to know whether the scribes’ versions of the languages they were writing had a lasting impact on regional varieties of the TLs. They are certainly contact-induced innovations, and they are certainly potential changes in the language; whether that potential for change was fulfilled is not especially relevant if our main interest is in understanding the range of contact phenomena in ancient as well as in modern times.

4. CONCLUSION. In spite of the considerable difficulties involved in the investigation of ancient language contacts, the future for this area of research is bright. Studies of ancient language contact are proliferating, as evidenced by the excellent papers presented at the conference from which this volume emerged and by the other articles and books cited above. It is inevitable, in the analysis of ancient languages, that our sources of information are too
limited to provide answers to all the questions we would like to ask; this is as true of the study of language contact in the ancient world as it is of other areas of linguistic investigation. But it is becoming increasingly obvious that studying ancient contact situations can lead to convincing arguments for contact-induced changes, including even contact-induced structural interference in cases where there is little or no lexical evidence to support the hypothesis.

Of course it is also true that efforts to argue for hypotheses of contact-induced change will sometimes fail, just as efforts to draw other kinds of historical conclusions sometimes fail for want of solid evidence. But if we exercise due methodological rigor in testing hypotheses in this domain, we can at least be confident that we can show the difference between a well-supported hypothesis and a truly inconclusive result.

REFERENCES


Emeneau, Murray B. 1956. India as a linguistic area. Language 32:3-16.


Fernández Marcos, Natalio. 2004. From contact to replacement: the first Bible translations. This volume.

Herodotus. 5th century BCE. The Persian wars.


Khan, Geoffrey. 2004. Aramaic and the impact of languages in contact with it through the ages. This volume.


Márquez Rowe, Ignacio. 2004. Akkadian as a contact language. This volume.


Reintges, Chris. 2004. Coptic Egyptian as a bilingual language variety. This volume.


Thurgood, Graham. 1999. From ancient Cham to modern dialects: two thousand years of language contact and change, with an appendix of Chamic reconstructions and loanwords. (Oceanic Linguistics Special Publication 28.) Honolulu: University of Hawai‘i Press.


