PROCEEDINGS OF THE TWENTY-FIFTH ANNUAL MEETING

OF THE

BERKELEY LINGUISTICS SOCIETY

February 12-15, 1999

SPECIAL SESSION

ON

CAUCASIAN, DRAVIDIAN, AND TURKIC LINGUISTICS

edited by

Jeff Good

Alan C.L. Yu

Berkeley Linguistics Society

Berkeley, California, USA

McCarthy, John and Alan Prince. (1995). Faithfulness and reduplicative identity. In *University of Massachusetts occasional papers in linguistics 18: Papers in Optimality Theory*, pp. 249-384. Amherst, MA: GLSA.

Nedjalkov, Igor. (1997). *Evenki*. New York: Routledge.

Prince, Alan, and Paul Smolensky. (1993). Optimality theory: Constraint interaction in generative grammar. *Technical Report #2 of the Rutgers Center for Cognitive Science*, Rutgers University.

Steriade, Donca. (1995). Underspecification and markedness. In John A. Goldsmith (ed.), *The handbook of phonological theory*. Cambridge, MA: Blackwell.

Stevens, Kenneth, S. J. Keyser and H. Kawasaki. (1986). Toward a phonetic and phonological theory of redundant features. In J. Perkell and D. Klatt (eds.), *Invariance and variability in speech processes*, pp. 426-449. Hillsdale, NJ: Lawrence Erlbaum.

Wu, Chaolu. (1996). *Daur*. Languages of the World/Materials vol. 93. München: LINCOM EUROPA.

Attractiveness and relatedness: Notes on Turkic language contacts

LARS JOHANSON
University of Mainz, Germany

0. Introduction

The aim of the present paper is to give a short account of some thoughts about possible ways of argumentation concerning the external relations of the Turkic languages, in particular the putative language family commonly referred to as 'Altaic'.

For the sake of brevity and clarity, I will dispense here with references to the vast literature on the subject and just refer to some previous work in which I have argued for the theoretical claims in more detail and on the basis of fuller empirical data.

1. Relatedness versus copying

Comparative linguistics disposes of a strong apparatus for judgments concerning cognate items on the basis of regular phonetic correspondences. However, the theoretical foundations for determining items as copied—or, with a traditional but inadequate term, "borrowed"—from other languages have so far been relatively weak. It will be argued here that stricter criteria are also required for the assessment of contact-induced similarities between languages.

The situation may be illustrated with the external relations of the Turkic languages (Johanson and Csató 1998), which exhibit indubitable similarities with Mongolic, Tungusic, Korean, and even Japanese. Historical linguists who do not believe in chance, i.e. accidental similarity, are confronted with two options in this case: to ascribe the parallels to *genetic relatedness*, or to *copying as a result of contact*.

2. The Altaic hypothesis

As regards the first option, the so-called "Altaic hypothesis" is still a matter of controversy. The existence of an Altaic protolanguage has not yet been proved unequivocally with the methods of comparative linguistics. This means that the Altaic languages have not been shown to be interrelated in the same sense as, for instance, Indo-European ones. This is even true of the relationship between Turkic and Mongolic. The difficulties are due to the enormous time dimensions involved in the reconstruction of the protolanguage in question and the poor documentary resources for a deep perspective of this kind, namely lack of older data, textual evidence, and significantly older written records of the languages concerned (Johanson 1990).

The second option means that the similarities are due to copying. It is beyond question that some of the languages involved, whether genetically related or not, have influenced each other for many centuries. Irrespective of their origin, immediate neighbors frequently share traces of areal interaction. It is often highly difficult to distinguish areal convergence from genetically conditioned divergence (Johanson, in print).

The advocates of the Altaic hypothesis, the so-called Pro-Altaicists, claim that the parallels are established on a set of regular phonetic and semantic

correspondences. They take the genetic relationship to be proven if connections of this kind can be demonstrated for a significant portion of the basic vocabulary.

The opponents of the Altaic hypothesis, the so-called Anti-Altaicists, explain the parallels as the result of copying. Pro-Altaicists answer that, in order to disprove genetic relatedness, it would indeed be necessary to demonstrate that the parallels are random and that there are no regular phonetic and semantic correspondences between the given languages. Anti-Altaicists maintain that it is also possible to establish regular correspondences between unrelated languages in cases of massive borrowing. Pro-Altaicists respond that massive copying of basic vocabulary is implausible.

The disagreement is understandable. Genetic relatedness cannot be observed directly, just the traces it may have left in the data available to the linguist. Different positions concerning the interpretation of such data are natural. For some reason, this is nevertheless a highly emotional domain (Johanson 1999c). Pro- and Anti-Altaicists are permanently at daggers drawn with each other. The discussion is full of sharp polemics, use of military terms, strict assignment to hostile camps, emotional attacks, hints at psychological roots of the controversy, etc. Being an Anti- and a Pro-Altaicist is equally blamable in the opposite groups. And agnostic positions are condemned in both camps. In spite of all the dangers in this minefield, a few comments on the genetic question will be ventured here.

Opponents of the Altaic hypothesis frequently urge its advocates to apply more stringent criteria. It would, however, be equally appropriate to direct this request to scholars who deal with copying in a rather loose way (Johanson 1995b). It is often observed that similar items that cannot be proven as cognates are declared "borrowings" without reference to any criteria whatsoever. Stringency is also needed in hypothesizing about copying processes. In order to find arguments for genetic relationship and contact, evidence for cognates and copies, it is advisable to utilize data from earlier research in historical linguistics and to pay attention to the findings of studies in language acquisition and psycholinguistics. Assumptions about correspondences due to copying should be tested in terms of empirical plausibility and diachronic naturalness.

3. Code-Copying

In the following, reference will be made to the so-called Code-Copying Model, an integrated descriptive framework for typological analysis of various contact-induced phenomena conceived of as linguistic 'copies' (Johanson 1992, 1993a, 1998, 1999b). The term *code-copying* is employed for the common interaction of linguistic codes mostly referred to as "borrowing". The crucial idea is that copies of elements from a foreign *model code* are inserted into a *basic code*, which provides the *morphosyntactic frame* for the insertion.

The choice of the term *copying* instead of *borrowing* is motivated by the important distinction between originals and copies. Copies are always part of the copying system and to some degree adapted to it; they are thus never identical with their originals. Code-copying always affects the structural characteristics of the basic code.

There are two kinds of insertion depending on the assignment of the two codes: *adoption* and *imposition*. In the case of adoption, speakers 'take over' copies from a secondary language into their primary language. In the case of imposition, speakers 'carry over' copies from their primary language into their

variety of a secondary language. The effects of adoption and imposition may be rather different.

In the case of *global copying*, a foreign item with all its properties (meaning, shape, combinability, or frequency) is the object of the copying. In the case of *selective copying*, only individual aspects of a foreign model are copied: meaning, shape, combinability, frequency (see, e.g., Johanson 1993b). Only global copying yields form-and-function items suitable as objects of comparison in genetic studies, i.e. as evidence against cognateness; e.g. the Tatar lexeme *qara-* 'to look' and the Tuvan verbal noun suffix -*(I)lda* are both globally copied from Mongolic.

4. Turkic language contacts

The Code-Copying Model has been used to describe typological tendencies in language contacts, e.g. how Turkic languages have influenced others and vice versa. This is a vast research area, since Turkic has been in close contact relations with numerous languages throughout its history. Strong mutual influences with Indo-European have arisen through a long-lasting Turkic-Iranian symbiosis in Central Asia. Other contact partners include Mongolic, Slavic, and Finno-Ugric varieties. The code-copying framework may serve to describe both ongoing contact phenomena such as the development of Turkish in Germany and processes that can only be observed in a diachronic perspective, e.g. Slavic influence on the Turkic Karaim language during its history of at least 600 years (see Csató, this volume).

There is abundant material to study the results of all these contacts. Some of them have been presented in a monograph on Turkic language contacts (Johanson 1992; English version forthcoming). The book tries to sum up what kinds of contact-induced changes have affected the structures of Turkic languages and their non-Turkic counterparts, what structural factors have been decisive for copying, what kinds of non-lexical items and properties have been copied under various social circumstances, etc.

5. Changes in the morphosyntactic frame

The study in question suggests that almost any feature can be copied given the appropriate social conditions. The morphosyntactic frame provided by the basic code may itself undergo essential changes through copying of new morphosyntactic features, e.g. markers of grammatical functions. Under conditions of sustained intensive contact, speakers may progressively restructure their basic code on the model of a dominant code. Successive frame changes pave the way for the insertion of further grammatical copies (Johanson 1999a).

This is sufficiently documented in the history of Turkic. The Turkic group includes several high-copying languages that display a good deal of frame innovation due to copying from genetically unrelated and typologically different languages. Some go very far in altering their basic morphosyntactic frame. Due to long and intensive contact, Turkic varieties spoken in Central Asia and Iran have developed considerable similarities with Persian. Gagauz, predominantly spoken in southern Moldova and neighboring parts of Ukraine, has developed under strong Slavic influence. Karaim, now spoken in Lithuania and Ukraine, has also converged with Slavic contact languages. Salar, spoken in China, has copied numerous frame-changing elements due to sustained contact with Chinese and Tibetan. The same is true of some languages that have been heavily influenced by Turkic.

Note that such changes in the basic code have taken place without a changeover to the model code. Heavy copying may make the basic code very similar to the model code without leading to a code shift. The influenced languages have not been transformed into the languages that have influenced them. Thus, Ottoman Turkish remained Turkic irrespective of its overwhelming load of copies from Arabic and Persian. It copied practically everything except a few basic function markers. Karaim has remained Turkic in spite of all its frame-changing developments. High-copying codes remain identifiable by the non-copied core elements of the frame. Even if the consecutive changes a given language has undergone may have led to considerable deviations from the features typical of its genetic group, it may still be classified with that group.

6. Attractiveness

Evidence from contact linguistics may allow hypotheses about what parts of grammar are mostly affected in copying processes. The study mentioned above (Johanson 1992) discusses what makes structures relatively 'attractive', more susceptible to copying, more readily copiable than other features.

Most Turkic languages today exhibit attractive features in the sense of transparency and regularity. Throughout their history they have, at least in the central areas, tended to abolish and change "marked" structures, to promote salient semantic and material structures, to regularize paradigms, to substitute attractive features for unattractive ones. The dominant Turkic type of today is a result of continued reinforcement of regular and transparent structures. Even the earliest Turkic known to us—documented in 8th century inscriptions—may be the result of a leveling koineization.

The study of well-known contact-linguistic cases suggests that one type of circumstantial evidence in the discussion of the Altaic hypothesis may be based on the concept of attractiveness. It might be used as an argument to strengthen or weaken the probability of an element having spread as the result of contact. Evidence for relatedness might be premised on features that have proved unattractive for copying. Empirical knowledge of what items are less likely to be copied may lead to hypotheses about prehistorical copiability.

7. The least copiable non-lexemic items

One crucial issue concerns the copiabililty of free and bound non-lexemic items. In order to gain insights of relevance for the questions of relatedness it should be determined empirically what morphological elements are most resistant to contact-induced frame changes. What elements refuse being replaced by copies in the frame-providing code? What native items are finally left intact even in high-copying languages? What items permit us to classify a language with a particular genetic group even if it strongly deviates from the features typical of that group?

Proponents of the Altaic relationship have identified a fair number of common morphological elements as putative cognates. But they have also characterized numerous unintelligible elements of word forms as 'suffixes', without accounting for their functions. Anti-Altaicists, on the other hand, almost never acknowledge any constraints on copying.

Copiable items of this kind often exhibit a salient semantic and material structure. Morphological opacity and irregularity are unattractive properties. The items copied have a relatively specific meaning and frequently a relatively elaborated shape. This may mean that they are mainly copied at early stages of

grammaticalization. Thus, relators such as postpositions are easily copied while they are still comparatively salient. The ones based on nouns belong to the most copiable relators; e.g. Turkish *sebebiyle* 'because of' is based on Arabic *sabab* 'reason' (see Johanson 1993c, 1996a). A further case in point is the high copiability of discourse-relevant conjunctions (Johanson 1997, 1998). Items that have arrived at the end of their grammaticalization path—exhibiting less salient structures, more reduced shapes, and more general, abstract meanings—seem to be less suitable candidates for global copying.

If we find corresponding items of less copiable kinds, i.e. similarities between Altaic languages in domains that are empirically less susceptible to influence, this might be an argument for genetic relatedness. Items that can be shown to be old and subject to long grammaticalization processes, provide particularly good arguments.

It may, however, be very difficult to distinguish between mere lookalikes and true cognates in this domain. The situation is never so simple that two putative cognates "look the same" and "mean the same". In the development of suffixes, otherwise valid phonological laws are often violated, and exceptional developments may be expected. Irregularities in bound morphemes are often removed through analogy, which may render reconstruction and proof of cognateness impossible. Items etymologically connected with each other may be at different stages of grammaticalization. High-copying languages possess several historical layers of copies reflecting different contacts. Each layer makes it increasingly difficult to distinguish between non-copied items and nativized copies.

8. Candidates for cognateness

Certain categories have proved less copiable in the documented history of Turkic.

(i) Relators such as case suffixes, in particular markers of central syntactic functions as the accusative, dative and genitive suffixes, are all of high age and display highly generalized meanings as well as less salient shapes. There are no known examples of case suffixes copied into Turkic varieties. However, as a result of imposition (see section 3) due to very long and intensive contact, Turkic case markers have been globally copied into North Tajik dialects.

(ii) Bound aspect-mood-tense markers also have abstract, generalized meanings. In the course of their functional development, they have often been replaced by other native items which have renewed the expression of their previous functions. However, there are no known examples of items globally copied from non-Turkic models.

(iii) Personal pronouns—as well as predicative suffixes going back to them—have been rather stable in most Turkic languages. There are no known examples of copying.

(iv) Markers of actionality and voice filling the position next to the primary verb stem have not been replaced by copies from other codes. Some of them have, however, vanished, and some have been replaced by periphrastic postverb constructions.

There are also other examples of native elements that have remained stable in the history of Turkic. Thus, copular verbs ('to be') and ancillar verbs (e.g. 'to do') have been replaced by new native verbs (e.g. *er- > tur-* 'to be', *qil- > eyle- > et- > yap-* 'to do'), but not by copied ones.

Looking at the categories i-iv in a comparative Turkic-Mongolic perspective, we may observe the following:

(i) Certain case suffixes show similarities, e.g. locatives such as Mongolic *-dA* ~ vs. Turkic *-DA*, the Mongolic terminative in *-čA* vs. the Turkic equative in *-čA*. The explanation of the Ordos, Khalkha, and Kalmyk accusative suffixes (type *-IG*) as global copies of the Orkhon Turkic accusative suffix in *-(°)G* seems little convincing.

(ii) Similarities are also observed between certain aspect-mood-tense markers, which all have reduced shapes and must be of considerable age, e.g. the Mongolic habitual in *-dAg* vs. the Turkic participle in *-DOQ*. It is even possible that a finite predicative marker such as Mongolic *-ba* (*irebe* 'has come') may correspond to a converb marker such as Turkic *-b* (e.g. *kelib* 'having come') (Johanson 1995a, 1996b).

(iii) Pronouns display similarities, e.g. Mongolic *bi* vs. Turkic *ben* 'I'. The oblique stems contain a so-called pronominal *n*, e.g. Mong. *minu < binu* 'my'.

(iv) There are similarities between the old actionality and voice markers that fill the position next to the primary verb stem, e.g. voice markers such as causative suffixes. The Mongolic cooperative suffix *-lčA* may be connected with Turkic *-(°)š*.

9. Irregularities

Interestingly enough, several of the cases just mentioned are exactly the ones in which Turkic has preserved traces of older irregularities. The predominantly regular and transparent morphological structures of contemporary Turkic languages evoke the deceptive impression that they have no secrets to reveal. On closer inspection, we find cases of morphological opacity and irregularity that appear typologically inconsistent with the rest of the structure. Anyone familiar with modern Turkish will know phonologically unpredictable allomorphs of personal pronouns, of causative suffixes, and of the old present tense marker in *-r* ("aorist"). Personal pronouns display vowel mutation (Turkish *ben* 'I', *bana* 'to me', Early Turkic *ben* 'I', *bini* 'me'), a phenomenon at variance with the agglutinative language type.

These are less attractive features of older stages of development that have not been subject to regularization. Turkic has preserved them exactly in cases where we, on independent grounds, expect the lowest susceptibility to copying, and where we also find the most striking Altaic parallels.

10. Conclusion

Candidates for Altaic cognateness might well be searched for among the categories just mentioned. The similarities are too striking to be random; the parallels can hardly be accidental. While they do not offer conclusive proof of the Altaic hypothesis, they are indeed indicative of close and systematic relations. It is true that the functional parallels are not always quite clear and that there are also various phonetic problems which make it difficult to prove cognateness. On the other hand, it is difficult to think of these abstract and reduced elements as copied from foreign codes, unless the copying occurred at very early stages of grammaticalization. In any case, there is no reason to accept the uncritical assumption that grammatical items of this type are freely copied back-and-forth among unrelated languages.

Note that areal convergence is not, as has often been suggested in the literature, a valid argument against common genetic origin. Let us assume that Turkic and Mongolic are interrelated and that the earliest Turkic known to us (8th century) is a *koiné* with simplified, "attractive" structures. If this is the case, the earliest Mongolic known to us (13th century) can be expected to have preserved an older, more complicated stage of development. Later areal interaction has demonstrably led to considerable simplification in Mongolic. For example, the gender distinctions and the inclusive vs. exclusive opposition in the 1st person plural have more or less vanished.

Such cases of convergence between Turkic and Mongolic certainly do not prove conclusively that the language groups in question go back to a common ancestor that exhibited gender distinctions and an inclusive vs. exclusive opposition. It should, however, be equally clear that they do not disprove this possibility.

References

Csató, Éva Ágnes. (1999). Analyzing contact-induced phenomena in Karaim. *Proceedings of the special session of the 25th annual meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society.

Johanson, Lars. (1990). Zu den Grundfragen einer kritischen Altaistik. *Wiener Zeitschrift für die Kunde des Morgenlandes* 8, 103-124.

Johanson, Lars. (1992). *Strukturelle Faktoren in türkischen Sprachkontakten*. (Sitzungsberichte der Wissenschaftlichen Gesellschaft an der J. W. Goethe-Universität Frankfurt am Main, 29:5.) Stuttgart: Steiner. [English translation in preparation; London: Curzon.]

Johanson, Lars. (1993a). Code-copying in immigrant Turkish. Extra, Guus and Ludo Verhoeven (eds.) *Immigrant languages in Europe*. Clevedon, Philadelphia, Adelaide: Multilingual Matters, 197-221.

Johanson, Lars. (1993b). Zur Anpassung von Lehnelementen im Türkischen. Laut, Jens Peter and Klaus Röhrborn (eds.) *Sprach- und Kulturkontakte der türkischen Völker*. (Veröffentlichungen der Societas Uralo-Altaica, 37.) Wiesbaden: Harrassowitz, 93-102.

Johanson, Lars. (1993c). Typen türkischer Kausalsatzverbindungen. *Journal of Turkology* 2, 213-267.

Johanson, Lars. (1995a). On Turkic converb clauses. In: Haspelmath, Martin & König, Ekkehard (eds.) *Converbs in cross-linguistic perspective. Structure and meaning of adverbial verb forms -- adverbial participles, gerunds* (Empirical approaches to language typology, 13), Berlin, New York: Mouton de Gruyter, 313-347.

Johanson, Lars. (1995b). Wie entsteht ein türkisches Wort? Kellner, Barbara and Marek Stachowski (eds.) *Laut- und Wortgeschichte der Türksprachen* (Turcologica, 26), Wiesbaden: Harrassowitz, 97-121.

Johanson, Lars. (1996a). Kopierte Satzjunktoren im Türkischen. *Sprachtypologie und Universalienforschung* 49, 1-11.

Johanson, Lars. (1996b). Altaische Postterminalia. *Acta Orientalia Hungarica* 49, 257-276.

Johanson, Lars. (1997). Kopien russischer Konjunktionen in türkischen Sprachen. In: Huber, Dieter and Erika Worbs (eds.) *Ars transferendi. Sprache, Übersetzung, Interkulturalität*, Frankfurt: Peter Lang, 115-121.

Johanson, Lars. (1998). Code-copying in Irano-Turkic. *Language Sciences* 20, 325-337.

Johanson, Lars. (1999a). Frame-changing code-copying in immigrant varieties. In: Extra, Guus and Ludo Verhoeven (eds.) *Bilingualism and migration*, Berlin, New York: Mouton de Gruyter, 247-260.

Johanson, Lars. (1999b). The dynamics of code-copying in language encounters. Brendemoen, Bernt et al. (eds.) *Language encounters across time and space*, Oslo: Novus Press, 37-62.

Johanson, Lars. (1999c). Cognates and copies in Altaic verb derivation. Menges, Karl H. and Nelly Naumann (eds.) *Language and literature—Japanese and the other Altaic languages. Studies in honour of Roy Andrew Miller on his 75th birthday,* Wiesbaden: Harrassowitz, 1-13.

Johanson, Lars. (In print). Code-copying and linguistic convergence. Esch, Edith, Mari Jones and Agnes Lam (eds.) *Linguistic convergence*, London: Arnold Publishers.

Johanson, Lars and Éva Á. Csató (eds.) (1998). *The Turkic languages*, London, New York: Routledge.

Epenthesis-Driven Harmony in Turkish

ABIGAIL R. KAUN

Yale University

0. Introduction

Describing one particular language requires only a subset of the Optimality Theoretic (Prince and Smolensky 1993) constraints necessary for the characterization of all phonological phenomena in all languages. Nonetheless, it is generally assumed that all grammars contain the same (universal) constraints, and differ only with respect to how those constraints are ranked relative to one another. The model therefore entails that individual grammars contain numerous latent or inactive constraints. This paper provides evidence that constraints which never play an active role in determining optimal outputs in native vocabulary are nonetheless available to speakers and may be invoked when novel phonological contexts are encountered, as in the case of loanwords. Several studies have reached similar conclusions for Tagalog (Ross 1996), Finnish (Ringen and Heinämäki 1999) and Tuvan (Harrison, this volume).

Words borrowed into Turkish from foreign sources typically undergo epenthesis to break up an initial consonant cluster. For instance, *steno* is realized as *siteno*. These epenthetic vowels are always high, however their backness and rounding are contextually determined, as noted in Yavas (1980) as well as Clements and Sezer (1982). This paper reports on the findings of an experiment conducted to determine the range of harmonic patterning of such epenthetic vowels. The results indicate that as described in the earlier studies, regressive harmony does target epenthetic vowels. The harmony "facts" vary considerably from speaker to speaker, however, and within speakers it is frequently the case that more than one harmony pattern is judged to be acceptable for a given word. I present an account of the results of this experiment focusing in particular on cross-linguistic evidence for the existence of the relevant inactive constraints that subjects must recruit to accommodate the novel contexts introduced in loanwords.

1. Patterns of regressive rounding harmony

Regressive harmony targeting epenthetic vowels in loanwords is described in Yavas (1980) and Clements and Sezer (1982). Both demonstrate that the backness of an epenthetic vowel is contextually determined. Vowels as well as velar and lateral consonants serve as backness harmony triggers. Similarly, these studies also document regressive rounding harmony targeting epenthetic vowels. The focus of this paper is limited to the phenomenon of regressive rounding harmony (hereafter RRH).

Both Yavas (1980) and Clements and Sezer (1982) report RRH targeting epenthetic vowels, however the patterns described are somewhat different. For both studies, high rounded vowels consistently trigger rounding harmony, as shown in the examples in (1):