Phonetic Laws and Grammatical Categories[1]

M. B. Emeneau
University of California, Berkeley

The discovery and statement of phonetic laws were for long a most important part of historical grammar. Reconstruction of earlier stages of a related group of languages, a 'language family', is based on the assumption that the recent languages are continuants/descendants of one earlier language, which is no longer extant in spoken form (though in some instances it may have been recorded). In some instances there is evidence for intermediate stages between the earliest and the latest languages.

Despite all recent disclaimers, I think it is still true that in historical grammar the various language stages are stated in terms of what since the 30s of this century have been called 'phonemes'. Certainly, phonological units of some kind are still operated with by practitioners of diachronic analysis. It is these units by use of which forms at all historical levels are stated. It is still true too that 'phonemes change', however one may interpret this postulate. The statements of such changes are 'phonetic laws'.

Who invented the term is of no present concern. From 1876 on, the term and the slogan 'regular phonetic change' were the trademark of a group of German scholars who dealt with the relationships of the languages included in the Indo-European family. Phonetic laws had been stated before then, since the early years of the 19th century, and much detailed statement of such relationships had been established. In the 1870s and 1880s exact statement of the methodology involved was debated with enthusiasm and acrimony. The debate centered around (among other points) the doctrine developed by scholars called then and since the 'Neogrammarians', viz., phonetic laws operate without exceptions; if exceptions are found in the data, the conditions for such exceptions must be looked for, and so on until all the phonological changes have been stated in terms of their conditioning. The Neogrammarians recognized that interdialectal and interlanguage borrowing also must be taken into account as producing exceptions to the basic laws.

But one type of exception caused much trouble and the Neogrammarian solution has led to debate ever since. One of their contemporaries, Georg Curtius, claimed (as early as 1870) that some exceptions to regular phonetic change were conditioned by the grammatical categories

of the exceptional forms. To the Neogrammarians this type of conditioning was anathema (this invidious term is justified by the acrimony of the debate on the matter). Their position (as stated by Bloomfield 1933, p. 364) was that phonetic laws were 'purely phonetic process[es] ... neither favored nor impeded by the semantic character of the forms which happen to contain the phoneme'. Curtius had been analyzing classical Greek data. In Greek, IE *s is deleted between vowels, but remains after voiceless stops (and in some other conditioned positions); e.g. *snusó- 'daughter-in-law' : Skt. snuṣā́, Gk. nuós; *ǵeno/es- 'race, kind', genitive Skt. jánasas, Gk. géneos. But in the aorist -s- remains between vowels; e.g. etímēsa 'I honored', ephílēsa 'I loved'. Similarly also in the future, with suffix *-s-. IE *y likewise is deleted between vowels; the optative suffix *-oy- appears in légoimen 'we would say' and most other members of this paradigm, but also in légoien 'they would say'. Brugmann's final analysis in 1885 was that the 'irregular' appearance of -s- and -i- (< *-y-) was due to restoration based on forms in which these sounds were 'regularly' retained (e.g. égrapsa 'I wrote', épleksa 'I wove', égeusa 'I gave a taste' [< *egeus-sa; *-ss- > -s-]; légoimen). This is 'analogical restoration'. 'Analogy', whether resulting in restoration or innovation, became in the Neogrammarian theory the all-important twin of regular phonetic change. Curtius's attempt to add grammatical conditioning as an alternative to phonetic conditioning in the statement of phonetic laws became heterodoxy to Neogrammarians.[2]

If conditioning by grammatical category was driven underground, as, e.g., in the type of statement represented by Bloomfield as noted above, and continuing down to treatments by such scholars as Dyen, Hall, and Hockett in the 1960s, there seems nevertheless to have been something of a hankering for the Curtius position. Robins, e.g., wrote (1978, p. 3): 'in some respects Brugmann is perhaps not so far from Curtius as is sometimes thought; ... Curtius would make the grammatical functions of the /-s-/ and the /-i-/ the direct cause of their survival, while Brugmann argues that the retention of such sounds, which otherwise disappeared in intervocalic position, was due to analogic association with paradigmatically associated forms like /édeiksa/ and /légoimen/'. This seems like whistling in the wind: stressing the mention of 'grammatical functions'/ 'paradigmatic association' in the statement of the two positions hardly brings them close together so long as Brugmann insists on 'analogic association'.

We should note that this debate, with all its clean-

cut opposition, originated and was specifically argued in the milieu of Indo-European studies, as was stated by Collinge, who (1978, p. 84) wrote of 'practical rigour applied to copious detail in limited areas', the 'limited areas' being the IE languages. Nevertheless, Bloomfield used the Neogrammarian doctrine to great advantage in Algonquian studies, and scholars concerned with other language families have followed suit. But in Finno-Ugric, a language family that has been cultivated diachronically even a little longer than Indo-European, concerned scholars have apparently never adopted the Neogrammarian doctrine in the matter of demanding phonetic conditioning and neglecting grammatical categories as conditioning factors in phonetic change. Anttila (1972, p. 79), in a section headed 'Grammatical Conditioning of Sound Change' (pp. 77-84), says: 'Clear evidence for grammatical conditioning comes from Baltic Finnic and Lapp; and, in fact, Finno-Ugric scholars have always used such information, even while it was theoretically undesirable in the main stream of linguistic inquiry', which was of course Indo-European studies. He gives a very straightforward example, for which apparently no other explanation (e.g. analogical) is possible: 'in Karelian, word-final nasals have dropped unless the nasal is the sign -n "gen. sg.", as in venehe-n "of a boat"'; and there are other examples included in his discussion.

The most modern linguistic theory abandons the Neogrammarian limitation of the conditioning factors of change to purely phonetic ones. Most instructive (and probably the most authoritative) is Postal's 1968 treatment of the stand of generative grammar. In his Aspects of Phonological Theory (pp. 233-44), he presents a collection of quotations which set forth the Neogrammarian position on 'regular phonetic change', labeling the doctrine (rather confusingly) as that of 'modern linguistics', meaning apparently 'American structuralism'; the authors run from Bloomfield (1933) to Dyen, Hall, and Hockett in the 60s. He then (p. 240) distinguishes two views on regular sound change as regards 'the kinds of environments' in which they occur. '(PI) Autonomy' is the Neogrammarian view which he rejects: 'the view that no regular sound changes require reference to morphophonemic or superficial grammatical environments'; '(PII) Non-Autonomy' is 'the view that some regular phonetic changes take place in environments whose specification requires reference to nonphonetic morphophonemic and/or superficial grammatical structure' (the emphasis in the statements is his). This latter is Postal's own view,

which he clarifies at some length, saying: 'generative grammar, which recognizes many rules with nonphonetic environments and likewise claims that such rules may therefore be added over time and thus play a role in describing sound change'. Presumably, the phrase 'superficial grammatical structure/environments' means what we have been discussing.

Since the Neogrammarian view did not admit of non-phonetic conditioning as a determinant of phonetic change, there was little stimulus leading to a search for contrary examples. Indeed, the need to explain otherwise some possible cases of non-phonetic conditioning (e.g. the Greek aorist) led, so Postal thought (p. 334, fn. 3; though with some slight reservation), to the classification of some instances of regular phonological change as 'analogy'. He says: 'I think that the term "analogy" has been used very misleadingly to refer to cases of perfectly regular phonological change in which part of the conditioning environment involves Surface Constituent Structure; i.e. changes which happen only in nouns, or only in verb stems, etc.' Postal's lead was followed by King (1969, pp. 119-27), who used as example the classical Greek data already discussed. Anttila (1972, pp. 98-9) discussed the classical Greek future with its suffix -s-, and wrote: 'sometimes morphological iconicity is so strong that sound change does not enter at all', and suggests a possible statement for Greek: 'intervocalic s drops, unless it means "future"'; he added that the aorist and the dative plural (suffix -si) are to be included.

Other generative theorists have provided similar treatment. Kiparsky (1971, p. 609) writes: 'It has lomg been noted that sound changes can depend on grammatical categories. We can add that particular kinds of grammatical categories are going to block sound changes from taking place'; this is in a summary of 'conditioning factors', with no examples given. Relevant examples, as in Postal, King, etc., are sometimes opaque or even controversial-sounding. Postal's example from Mohawk (1968, pp. 245-54) seems convincing: the proto-Mohawk-Oneida *kw sequence, like a number of other consonant clusters, has in Mohawk an epenthetic [e] inserted, except when the cluster means '1st person plural'. Anttila and Kiparsky have drawn much material from the Finno-Ugric languages; one example from Anttila is given above.

Considering this sketch of the history of theoretical attitudes to grammatical categories as conditioning for phonetic change, it would seem that an explicit study of the typology of rules for change would be

desirable. This is not the object of this paper; the presentation of further examples of changes conditioned by membership in grammatical categories should be welcome, and I present several. They range from one case that cannot be explained otherwise, viz., Tamil expressives, to several for which in the past phonetic conditioning was provided of a very uncertain nature. For these the admission of grammatical category as the condition for exceptional phonetic change now seems most economical. One of these has to do with Indo-Aryan numerals, the other involves English forms that have as initial the voiced dental fricative ð.

Expressives, especially in Tamil

Once it becomes admissible to look for grammatical or semantic categorics as possible determinants of phonological change or structure, the data that are usually referred to as 'onomatopoetic' or 'symbolic' need no longer be skittishly treated as somehow marginal. Synchronically, structure will be found and often will be amenable to phonological statement. Hock 1986 (especially pp. 177-9 and 287-90), following after such treatment as that in Bloomfield (1933, pp. 244-6), has considerable discussion of English, German, and Hindi material. He starts with forms that have reference to various types of sound (p. 177; e.g. English clatter, clack, clang, clap, clash, crack, crash, splash, splatter, etc.) and of visual effects, especially of light (pp. 288-9; e.g. flicker, glimmer, glisten, glitter, etc.). He mentions 'other areas of vocabulary', and might have found much in English, e.g. his heart fluttered, she (a menopausal woman) had a hot flash (anglicè a hot flush). Most of Hock's treatment is synchronic, attempting to find statements of structure, though he makes tentative forays into the history of the English material.

There have been studies, both synchronic and diachronic, which have added much in a generalizing direction that Hock might have taken into account, though the prevailing diachronic interest of his work probably inhibited this somewhat. Diffloth 1976 made a very searching morphological and semantic study of data frpm Semai, an Austroasiatic language of Malaysia, which is rich in relevant material. In 1969 I had studied pertinent data from Kota, a Dravidian language of South India, and began a comparative study of parallel data in the Dravidian family, especially South Dravidian, including Tamil, and in Indo-Aryan; this was reprinted in 1980 (pp. 250-93), and some programmatic remarks were made in the same volume (pp. 6-9), suggesting that all (or most) of the languages of Southeast, East, and Central

Asia have elaborate structures of this kind. I had already in 1951 (pp. 159-200) made a collection from dictionaries, with preliminary analysis, of such formations in Vietnamese; Thompson (1965, pp. 154-72) treated the problem with a very different analytical methodology.

Study of the more exotic languages has already made it clear that the pertinent material usually belongs to a morphological-syntactic class that is distinct from other major classes of the language; this was found in Semai and in the Dravidian languages, the morphology being quite distinct from that of verbs and of nouns. In phonology it has sometimes been determined that the category has different rules from the rest of the language. This has been demonstrated for Semai. It has not yet been completely demonstrated for the Dravidian language Tamil, but there is clear evidence looking in that direction. E.g., though Tamil forms in general, nouns and verbs, apart from some borrowed Indo-Aryan forms, do not have voiced stops, retroflex consonants, or l in initial position, some of the 'onomatopoetic' forms do. A complete study of Tamil data is desirable; this must be undertaken with or by Tamil speakers,[3] since the Tamil writing system has no way of indicating voiced stops as contrasting with voiceless. Consequently, the otherwise very adequate Tamil Lexicon writes with initial voiceless stops those onomatopoetics which in the standard colloquial have initial voiced stops and which comparison with etyma in other languages of the family (such as Kannaḍa or Telugu) shows had initial voiced stops in reconstructed Proto-South-Dravidian or even Proto-Dravidian; e.g. (DEDR 2576; etyma with j- in Kannaḍa, Telugu, Pengo, Maṇḍa, Kui, Kuwi) Tamil jilujilu 'to feel chilly'.[4] It should be noted that the Tamil Lexicon records old literary occurrence (even in the earliest texts, the Sangam) for some of the pertinent data, but the use of the normal writing system in the whole of the literary record prevents recognition and notation of initial voiced stops in such old forms when they happen to be vouched for in the modern standard colloquial (or in dialect usage).

It is even possible that Hock's 'onomatopoetic verbs' of English, tabulated by him on p. 177, should be identified as a lexical class of distinctive semantic type, part of whose identification is phonological, statable in terms of initial and final combinations of consonants and consonant clusters.

The relatively impoverished systems that have been investigated among the European or Indo-European languages hardly give a glimpse of the richness of the

semantics of this category as it is met and recorded in some other languages (including in fact the Indo-Aryan languages). In my dealing with Dravidian and modern Indo-Aryan material (1969, p. 284 = 1980, p. 263) it was clear that terms relating to sounds and visual/light effects formed only a small part of the material. I ventured a semantic statement: 'the class denote[s] varied types of sensation, the impingement of the material world, outside or within the person, upon the senses--not merely the five conventionally identified senses, but all the feelings, both external and internal'. There were many types of sound and of light effects, but also many such items as 'sudden motion', 'trembling', 'irritation in the throat', 'heavy sensation as of bad digestion', 'mental agitation or confusion', 'smiling', 'a dog's snapping', etc., etc. Diffloth's Semai examples are even more exotic: 'appearance of a long line of people carrying heavy loads', 'continuous wavy lines seen in one glance', 'irregular flapping circular movements (as of a tortoise's feet struggling to escape)', 'appearance of nodding constantly', etc., etc. It becomes obvious that the usual terminology such as 'sound imitation', 'sound symbolism', 'onomatopoetic' has little universal value, and Diffloth's terminology (1976, pp. 263-4, fn. 2), of which 'expressives' labels the most inclusive class, is preferable; I adopted it in my later treatment (1980, p. 7) and use it from here on.

Hock (pp. 177-9) made a tentative start at historical treatment of the English 'sound' and 'light' words, finding that the most tempting solution was in terms of analogy. This seems to be a Neogrammarian avoidance of historical explanation of phonological change as conditioned by grammatical/semantic category. And yet it is now possible to cite Tamil expressive forms in which membership in this class blocks the Tamil historical palatalization rule. By this rule[5] Proto-Dravidian initial *k- becomes Tamil c- when it is followed by one of the front vowels *i, *ī, *e, *ē; the rule is blocked when a retroflex consonant follows the front vowel. The rule operates elsewhere than in the expressive class; there are exceptions (not too numerous), some to be found in words borrowed from Kannaḍa (in which there is no palatalization), some still requiring study. The blocking of the palatalization rule in expressives has already been stated by Annamalai 1968, but since he neglected the appearance of initial voiced stops in some of the expressives (writing, e.g., only k- and not g- in forms in which the voiced stop occurs), reexamination of his results is necessary (but cannot be undertaken here in detail except for those with velar stop initials). It is possible too that the palatalization

rule needs some slight addition involving alveolar consonants as a blocking factor in general; this requires detailed investigation elsewhere. It should be noted too that Telugu has a palatalization rule almost identical with that of Tamil, with the difference that the presence of a retroflex consonant (or perhaps also an alveolar consonant) following the front vowel does not have the blocking effect that it has in Tamil (and Malayalam). It has been suggested too (Krishnamurti 1961, p. 25, §1.59), because initial g- sometimes is found in Telugu instead of *k- palatalized to c-, that the initial voicing occurred before the onset of the palatalizing rule and that only *k- (and not g-) was affected. Some of the Tamil expressives have initial g- rather than k-; whether Krishnamurti's Telugu suggestion is needed for these Tamil unpalatalized forms with g- is not yet clear.

In spite of the uncertainties affecting some of the pertinent Tamil data, scrutiny of the data (in DEDR) finds no palatalization in any of the expressives which have etyma beginning with a velar stop--whether the expressives have initial k- or g-, and whether an alveolar consonant following the front vowel might or might not be a factor blocking palatalization in words that do not belong to the expressive category.

The data are as follows,[6] beginning with three items which, if they were words of other categories, would certainly have palatalization.

DEDR 1515 Tamil kiccukicc-eṉal, kiccukicc-eṉal 'chirping' : k- in Malayalam, Kannaḍa, Tulu, Telugu (kicakica).

DEDR 1516 Tamil kiccukiccu mūṭṭu 'to tickle (transitive)' : Toda kiskwïṯ- id., kiskwïḷ- 'there is tickling' (with dative).

DEDR 1594 Tamil kettukkett-eṉal, ketaṅketam-eṉal 'palpitation of the heart through fright' : Kannaḍa kettu 'to quiver, shiver, tremble', Tulu kittuni 'to flutter, be in agitation'. The Tamil forms are not known to my informant; the Tamil Lexicon records them as 'local', i.e. from some (unidentified) local dialect.

DEDR 1593 Tamil girugir-eṉal 'creaking (of door)' : Tulu kirikiri, girigiri, Telugu kirakira. The Tamil item is not in Tamil Lexicon or DEDR, but is given by Annamalai 1968, p. 16 (of course with k-).

DEDR 1595 Tamil girugir-eṉal 'giddiness' : g- in Kannaḍa, Tulu, Telugu.

DEDR 1575 kilukilu, gilugilu 'to rattle' (also as noun 'a child's rattle') : g- in Kannaḍa, Telugu.

Three expressives in which a retroflex consonant following a front vowel blocks palatalization as it would in words of other categories may be noted, since

in one, DEDR 1545, a Telugu etymon also retains its initial velar unpalatalized, suggesting that the rules given for expressives in Tamil may apply also to Telugu (complete study of the Telugu data has not yet been undertaken).[7]

DEDR 1545 Tamil giṉgiṉ-eṉal 'tinkling' : in Kannaḍa and Tulu k- and g-, Telugu kiṉakiṉal-āḍu 'to tinkle'.

DEDR 1530 Tamil giṯugiṯu 'to tremble, shiver' : an etymon only in Malayalam.

DEDR 1531 Tamil giṯugiṯu 'to rumble' : an etymon only in Malayalam.

The conditioning, then, for the blocking of the palatalization rule in Tamil is membership in the category of expressives, no phonological conditioning being possible.

Indo-Aryan numerals

It is notorious that in the modern Indo-Aryan languages (NIA = modern Indo-Aryan), except for the geographically separated Sinhalese, the numerals below 100 are extremely 'irregular', requiring for their description a complete listing of forms from 1 to 100 or a very complex set of generative statements. For their learning, whether for native speakers or foreign learners, a most unusual effort of memory is necessary. A glance at any grammar of Hindi, Bengali, Gujarati, Marathi, or any other language of the sub-family is enough to demonstrate this; an excellent modern synchronic study of the Hindi data is available in Bright 1972.

These numerals derive historically from a much simpler Old Indo-Aryan (OIA) / Sanskrit (Skt.) system, whose irregularities are no greater than those that are found in other Indo-European languages of the older period. Statement of the diachronic rules for derivation of the NIA numerals from those of OIA will be complex, but has not yet been completely attempted. Here we take for granted various 'regular' changes and only look at what is one of the most important of the 'irregular' rules that must be stated in going from OIA to Middle Indo-Aryan (MIA) in the teens, i.e. 11 to 18.[8] In OIA these are compounds of the units (1 to 8) plus daśa '10', with a very few statements needed about the shape of some of the units, viz., eka- '1' → ekā-, dva- '2' → dvā-, tri- '3' → trayō-, ṣaṣ- '6' (irregular in its inflection) → ṣō-[ḍaśa], aṣṭa- '8' → aṣṭā- (no change in catur- '4', pañca- '5', sapta- '7'); the OIA forms are: ekādaśá '11', dvādaśa '12', trayōdaśa '13', caturdaśa '14', pañcadaśa '15', ṣōḍaśa '16', saptadaśa '17', aṣṭādaśa '18'.

The regular rule for the representation of OIA inter-

vocalic -d- in MIA is that it is retained in Pali, and in the literary Prakrits (Pkt.) it is in general deleted (or in Jain writings replaced by a 'weak' ẏ); e.g. Skt. hṛdaya- 'heart', Pali hadaya-, Pkt. hiaa-, hiẏaẏa-; Skt. udaka- 'water', Pali udaka-, Pkt. uaa-. So far as I know, no special study has been made of what happens to -d- when it is initial in the second member of a compound; certainly there is not available any complete collection of data. Turner-Turner's 1971 publication of phonetic analysis of the Turner 1966 dictionary material (pp. 170-1, s.v. -d-) yields a very few pertinent items. These show the same development as in simplex words; e.g. CDIAL 4371 Skt. *grāma-dāra- 'village boy', Pali gāmadāraka- id., Pkt. gāmāra- 'rustic, fool' (< *gāma-āra-); CDIAL 7799a Skt. para-dēśa- 'foreign country', Pkt. paraēsa-; similarly in verbs (with initial d-) plus prefixes, and in noun derivatives therefrom; e.g. CDIAL 8668 Skt. pradēśa- 'a direction, region, country' (< pra-diś- 'to point out [e.g. a direction]'), Pali padēsa-, Pkt. paēsa-; CDIAL 7836 Skt. pari-dā- 'to give, grant', noun pari-dāna-, Pkt. pariẏāṇa- 'general gift'; CDIAL 11734-5 Skt. vi-dār- 'to tear to pieces', Pali vidār-, Pkt. viār- 'to split'.

In contrast with these regular developments, the teen numeral compounds in the literary Prakrits have intervocalic -d- of Sanskrit and Pali replaced by -r- (which remains down to the present in the NIA languages except for those of the far Northwest). There are differences in detail in the various Prakrits, especially in the first members of the compounds (the units), but also in that -ś- (of daśa) appears as -s- in most of MIA, but as -h- in Māhārāṣṭrī Prakrit (and in the latest MIA stage Apabhraṃśa and then in the modern languages): ĕkkārasa, ēgārasa, ēāraha '11' (Hindi: CDIAL 2485 igārah, Bright 1972 gyārah), bārasa, bāraha '12', tērasa, tēraha '13' (trayō- replaced by *trayē-), paṇṇarasa, paṇṇaraha '15', sattarasa, sattaraha, (also sattara) '17',[9] aṭṭhārasa, aṭṭhāraha, (also aṭṭhāra) '18'.

It should be noted that in the Prakrit inscriptions of Aśoka (c. 260 B.C.) the few instances of teens all have -d- and not -r-; so also in Gāndhārī Prakrit (Dhammapada and Niya). But in the Hāthīgumphā inscription (? 2nd century B.C.), there occurs paṃdarasa-vasāni '15 years', with -r- (for pañca > paṃda '5', cf. Hindi pandrah '15' and other forms in CDIAL 7662, and Woolner 1928, p. 206). In the Pali texts, as we have them, there is only -d- in these forms. These, as well as forms with -r-, are given by Geiger (1916, p. 103, §116.2) and Turner (1966, s.vv.). However, von Hinüber

(1986, pp. 167-8, §§400-2) reports that Aggavaṃsa's grammar Saddanīti (A.D. 1154; written in Burma) gives -r- forms (ēkārasa, etc.) for Pali; von Hinüber's interpretation is that such forms ('das echte mi. Wort') have been eliminated from the text manuscripts by Sanskritization ('Sanskritismus'), which, to be sure, is seen often enough in Pali (see, e.g. Emeneau 1978, p. 117, col. 1 top = 1988a, p. 158).

Alongside of these numerals which show an historical development -d- > -r-, there is another group of forms showing the same change and not definable in terms of phonetic conditioning. These are in Sanskrit bahuvrīhi compounds of various pronominal stems plus several noun derivatives of dṛś- 'to see', these nouns meaning 'appearance'; an example is the compound tādṛś- or tādṛśa- 'having that appearance', i.e. 'like that'. The pronouns that occur as first member of the compound are: tā- 'that', yā- 'which' (correlative of tā-), kī- 'which?', ī- 'this', ētā- id., anyā- 'other' ('like another, different'), mā- 'me', tvā- 'you sg.', asmā- 'us', yuṣmā- 'you pl.'. In addition, there is sadṛś-, etc. 'like, resembling' (with the comitative prefix sa-), which already in a Vedic ritual (śrautasūtra) text has become sādṛśa- (by analogy with tādṛśa-).

In MIA we find such forms as tārisa-[10] alongside of the regular tādisa-; the latest MIA (Apabhraṃśa) has taïsa, derived by regular deletion of -d-. For yādṛśa- Prakrit has yārisa- and jaïsa- (< *jādisa-); Gāndhārī Prakrit (Niya) has yadriśa- with -d- (-ri- < -ṛ-). For īdṛśa- the various Prakrits have īrisa-, īdisa-, īisa- (and also forms with ē- instead of ī-, thought to be analogically derived from demonstrative ēṣa-, etc.). And so on for other compounds. Caillat 1980 studied the whole problem in great detail, using the identifying term (Englished by me as) 'pronouns and adjectives denoting similarity'. It is clear from CDIAL and Caillat that the change d > r appeared earlier in these forms than in the numerals, since the Aśokan inscription at Girnar (in Kathiawar; c. 260 B.C.) has tārisa-, yārisa-, and ētārisa-, while the parallel forms in his other inscriptions at other localities have -d-. On the other hand, on the whole the modern languages do not have for the 'similarity' words forms derived from the MIA -r- forms, but rather forms derived from the MIA -d- forms with deletion of -d-. In the Pali texts -d- forms are the norm, but the grammarians, including Aggavaṃsa (as above for the numerals) and the earlier Kaccāyana (so Geiger 1916, p. 59, §43), teach also -r- forms (Caillat 1980, p. 36). But Caillat points out that in Pali verse, which in general is presumed to be

of some age ('de facture ancienne'), there are examples of sarisa- (< sadṛśa-) and sarikkha- (< sadṛkṣa-). In fact, as is easily seen from CDIAL 13118-20 and Caillat, the forms with sa-, which is not in origin a pronoun, have a different history than forms with a pronoun as first member: sari, etc. 'like' with -r- are the only forms that are found in NIA (except for Sinhalese, which has s derived from a -d- form).[11]

Historical explanation of d > r in these two classes of words, the teen numerals and the 'pronouns and adjectives denoting similarity', has not often been attempted. Turner (1937/1975) included the numerals in an article in which his chief concern was to investigate Indo-Aryan data in which phonetic changes that later were normal appear sporadically at an earlier period. This may well be the correct explanation for some of his data. Concerning the instances of our present concern, he makes a statement (1975, p. 367) about a 'tendency for an intervocalic dental to become a spirant [he intends ð] resulting in r'. However, since -d- did not become -r- later or elsewhere in the Indic material,[12] we can only find that Turner's general thesis has no real pertinence for our present data. His statement made in passing (1975, p. 361) that 'gabbling in the recitation of series of numbers has had some influence in producing such abnormalities' is of some interest, but has no relevance for the 'similarity' words.

Otherwise, Berger (1955, pp. 42-4; following Meillet and Jules Bloch [see Caillat 1980, p. 36, fn. 14]) attempts an historical solution on Neogrammarian lines. He posits dissimilation of -d- to -r- in forms that have a preceding dental, viz., tādṛśa- and tvādṛśa- yield tārisa-, and among the numerals saptadaśa yields sattarasa (presumably through *sattadasa), dvādaśa yields bārasa through an intermediate *dvārasa, and trayodaśa yields terasa through an intermediate *trēdasa/*trērasa. On the basis of these forms he derives all the other forms by analogy.[13] This explanation is summarized/followed by Caillat (1980, pp. 35-6), and also by Turner in his posthumous addition to CDIAL (1985, p. 48, no. 5760; following Caillat). Berger attempts no phonological explanation of r from d, such as Turner ventured. Von Hinüber (1986, p. 167, §400) sums up his report of these recent treatments with the phrase 'eine nicht befriedigend erklärte Lautentwicklung'.

Instead of this complicated use of analogy following dissimilation, the latter in itself being an irregular and sporadic process and not otherwise exemplified (at least in CDIAL), I prefer the type of conditioning that I discuss in this paper, that by membership in a gram-

matical/semantic category, in this case two categories, viz., the teen numerals and Caillat's 'pronouns and adjectives denoting similarity'. If this sort of explanation is admissible, it is certainly simpler than Berger's.[14]

English words with initial

Another set of data which, being conspicuous, begs for explanation but has not yet received any completely satisfactory solution, comes from English. We know from Grimm's Law that Indo-European (IE) initial *t- should appear in modern English as þ (spelled th); so it does in such words as thin (: IE *ten-, *tn̥-), thatch (: IE *teg-, *tog-), thigh (: IE *tewk-), thirst (: IE *ters-, *tr̥s-), thorn (: IE *trn̥-), thorough (: IE *ter-, *tr̥-). On the other hand, a considerable group of words spelled with initial th preceding a vowel,[15] and of the same IE origin (*t-) is pronounced in modern English with ð, i.e. the voiced fricative corresponding to voiceless þ. This group is more easily listed than defined. The list follows, with some of the obvious inflectional forms and derivatives and compounds enclosed in parentheses: thou (thee, thy, thine); they (them, their, theirs); this (these), that (those); the; there (thereafter, therein, thereto, therefrom, thereby, thereof, therefore ['for (= because of) that'], etc.), thence (thenceforth), thither; then; thus; than; though (although); beside that, a demonstrative pronoun and adjective, we should mention that in its use as a relative pronoun (the man that we saw) and in its use as a conjunction (I said that he was here); 'empty' use of there in such sentences as there was a man that I knew may also be mentioned.

That ð and þ are contrasting phonemes in English hardly needs demonstration, even though minimal contrasts are rare; cf. thy : thigh, thou : thou[sand], teeth : teethe, mouth (noun) : mouth (verb).

The historical origin in IE terms being what it is, a difference in the conditioning to produce a split and to yield the two results has of course been looked for. The scholarly literature is clear that no other Germanic language has anything comparable to this English split in the development of IE initial *t-. The chronology of the development in English is not nearly so clear, and in fact the only really certain evidence that I have found in the reference books (see Dobson 1957, pp. 2-6, 936) is that found in a Welsh transcription of a Hymn to the Virgin in early modern English. The spelling conventions used in writing Welsh are applied to the English text, and in those words of our list that

occur in the text (their, there, this, these, they, the, thee are listed by Dobson) the spelling is almost always dd- (i.e. ð-). The chronology of the Welsh documents is not exact, resting on a number of undated manuscripts. Dobson (p. 3) says that we have 'English pronunciation ... as it sounded to a Welshman about, or rather somewhat before, 1500'; elsewhere (p. 936) he says that the fourteenth century was when the voicing of the fricative occurred. It seems clear also that in Old English the initial þ was not voiced. I have found no more exact statement of the chronology than this.

Dobson and others, e.g. Sweet (1888, p. 191, §730), Jordan (1934, p. 183, §207), Moore-Marckwardt (1966, p. 139, §124.1), and Hock (1986, p. 87, §5.3.1, and pp. 48f., §3.6), attribute the exceptional development to occurrence in unstressed contexts of certain of the words, especially 'this, that, the, etc.' (Hock, p. 87, calling them 'clitics'), or 'the, that, then, though' (Sweet, p. 191), or 'thou, thee, this, that, then, than, thither' (Jordan, 'in unbetonten Worten'), or 'the weak forms of the words this, that, then, etc.' (Dobson, p. 936), and then from these by analogy in all the occurrences and in all the rest of the list. This might be so, but more exact statement would seem to be called for. Of the items in our list the only ones that could justly be characterized as usually unaccented or clitic are the article the, the conjunction than, that when it is a relative pronoun or a conjunction, and there in its 'empty' initial use. For all the others it is difficult to construct examples in which they are not accented somewhat, but usually rather heavily, with the possible exception of the 2nd person and the 3rd person pronouns subject and non-subject (thou, thee, they, them). The four forms that are usually unaccented seem a slight basis for such extensive analogical blending (contamination, or whatever it is to be called), especially since the historical connections of some of the forms must have been as unclear or unknown to 14th century speakers as they are to modern speakers.[16]

It is notable too that others have shied away from this explanation, usually in favor of identifying all our forms as a 'class', membership in which conditions the exceptional change. So Jespersen in 1891; in Selected Writings of Otto Jespersen (pp. 593-4) he lists most of the forms with a statement of the change;[17] he says nothing more, except to attempt to date the change, and so really only identifies the class. Turner (1928/1975) in a wide-ranging collection of exceptional phonetic developments in pronouns in many languages finally mentions the English data (1975, p. 317), listing 'the,

that, this, then, there, though, etc.'. He then concludes: 'The above examples show that we may reasonably ascribe a peculiar treatment to the initial of a pronoun'. Whether or not we may judge this a 'universal', we may at least note that Turner was willing to ascribe the English development to membership in a grammatical category.

Whorf (1956, p. 76) wrote of the occurrence of English ð initially only in the class (he called it a 'cryptotype') of 'demonstrative particles (the, this, there, than, etc.)', omitting the 2nd person singular pronoun. This undated synchronic discussion had not appeared in print earlier, but in 1945 he had published a paper on 'grammatical categories' which looked in the direction we are interested in (but only synchronically). Lehmann (1962, p. 72), in a section on writing, uses the example this in saying: 'the restriction of /ð/ primarily to a small set of morphemes similar in use, and to initial position in these'. He has nothing further, and his use of 'primarily' gives one pause as to whether he intended our whole class. Bloomfield (Morton)-Newmark (1965, p. 293) write: '[after 1600] initial /θ/ was replaced by /ð/ in a number of common words (all "pointing" words) like the, they, that, this, there, and so forth'. Disregarding the faulty late chronology, we may note the outright identification of membership in the class as the condition for the change.

Finally, I think we may regard as somewhat ambivalent or ambiguous Leonard Bloomfield's statement (only synchronic, to be sure; 1933, p. 147): 'Vocal gestures ... Less striking deviations from the phonetic pattern sometimes occur in words whose meaning resembles that of a pointing gesture. In English the initial phoneme [ð] occurs only in words of demonstrative and related meanings, such as this, that, the, then, there, though'. Thus to shunt off this class of words (neglecting also the 2nd person singular pronoun) to a position outside the normal structure of the language, allowed Bloomfield to ignore the historical problem involved (as I think he did).

It remains to find a term for this class. To point out that it contains all the pronouns, personal,[18] relative, and demonstrative (including demonstrative adjectives) that begin with a dental fricative, plus the definite article, and a number of 'pronominally derived function words', viz., adverbs and conjunctions (than, though, that), is an attempt to class all the list as somehow 'pronominal'. That this is an attempt to motivate a fairly late change by an etymological connection in a much more distant past seems obvious.

That than and though are etymologically connected with the demonstratives is true enough, but can the speakers who made the change have had any efficient sense of this connection? It is tempting to think that a simpler solution would be to define the class negatively as 'all words of this phonological shape (having initial dental fricative followed by a vowel) and not being nouns, adjectives, or verbs'. Whether this is acceptable I leave to others to determine within their theoretical systems.[19] But I think I have pointed to a prime instance where membership in a grammatical category is the condition for a sound change.

In discussion after the delivery of this paper, it has been pointed out to me[20] that my statement above, that 'no other Germanic language has anything comparable to this English split in the development of IE initial *t-', is incorrect. In modern continental Scandinavian (NContSc.; i.e. Danish, Swedish, Norwegian) there remain as such the initial fricatives f- and h- that result in Old Germanic by Grimm's Law from IE voiceless stops in labial and guttural position. The IE voiceless dental stop *t-, however, having become the fricative þ- in Old Germanic, results in NContSc. in the stops t- and d-. This split is entirely parallel in its incidence to that seen in modern English þ- and ð-. Problems are raised by the Scandinavian phenomena, e.g. why the dental stop differs in its development in this position from the other stops. Chronologically, the English and the NContSc. developments take place at slightly different periods. Haugen (1976, p. 266) summarizes the Scandinavian chronology: 'þ > t in Da[nish] by 1300, in Sw[edish] by 1400, in Nw [i.e. Norwegian] by 1450', without giving dates for þ > d in 'words that were frequently unstressed'. This dating seems late for Scandinavian, as spoken in northern England, to have had any influence on the English development, which in any case yielded quite different results phonetically. But, be that as it may, these two Germanic languages show the same split according to grammatical/semantic category. Whether a unified treatment is possible (? parallel 'Sapirian drift' in fairly closely related languages), remains a problem.

Notes

1. For material, discussion, and bibliography I must record my gratitude especially to Mrs. Kausalya Hart and Professors B. A. van Nooten, Herbert Penzl, and Gary Holland. They are of course not responsible for my errors or misjudgments.

2. If I do not explore Hugo Schuchardt's writings on this subject, it is largely because I am sure that he said nothing about Tamil expressives or Indo-Aryan

numerals. He may have discussed the English words with initial voiced dental fricative, but none of the modern treatments of this topic refers to him.
3. Through the kindness and expertise of Mrs. Kausalya Hart I have been able to check this point for a small section of the Tamil material in her dialect, which is the standard colloquial Tamil as spoken by the middle range of castes in Madurai (I think this is a correct statement; for the dialect problem, see Emeneau 1988b, p. 243, §6.3, and literature quoted there).
4. Tamil has a symbol for j different from that for c, found in a very few words, mostly borrowed (it is called a 'northern letter', as if used only for Indo-Aryan borrowings); for this item the Tamil Lexicon writes both, but my informant uses only j-.
5. The rule operates also in Malayalam, for which the blocking in the corresponding expressive category has not yet been investigated.
6. DEDR 1590 is not known to my informant.
7. DEDR 1597 Telugu ceṟalceṟal-āḍu 'to be angry or furious' and ciṟaciṟa 'sullenness, ill-humor' would seem to go against the blocking rule, but the items in entries 1597 and 1961 may not have been sorted correctly.
8. Skt. nava-daśa '19' is almost completely replaced in MIA by forms that are literally '20 less 1'.
9. I would have to invoke analogy with the forms for '17' to explain Pkt. sattarim̥ '70' (< Skt. saptati; in NIA, Hindi sattar, etc.). See fn. 13 below.
10. The old suggestion that tārisa- is cognate with Greek tēlíkos has been rejected finally, presumably because of MIA coexistence of -d- and -r- forms; see Caillat 1980, p. 33, Berger 1955, p. 44, and Mayrhofer 1956, p. 494, s.v. tādṛ́k.
11. For borrowing of sari in the Dravidian languages and the rich development there, see Emeneau-Burrow 1962, p. 31, entry 147.
12. A very few (sporadic) examples have been collected by Pischel (1900, p. 172, §245) and by Brough (1962, pp. 255-6, no. 259). In the present context, they must still remain exceptions requiring explanation.
13. In connection with this explanation he uses the word 'Verhaspeln', with which cf. Turner's 'gabbling' quoted above. —Berger also thus derives sattarim̥ '70', beside sattati, from saptati, by dissimilation. See fn. 9 above.
14. Hockett 1958, p. 258, writes: 'Numbers higher than one have some substitute-like properties and are perhaps to be so classed'; he does not proceed further.
15. 'Preceding a vowel' must be specified, since th- preceding r or w is always voiceless.

16. Verner's Law has several times been mentioned as somehow similar to this change. However, in detail the two are different (the position of the phonemes with reference to the accent—and what of the nature of the accent in the two changes?), and the chronological gap is great. This suggestion seems not very useful.
17. Pp. 579-616 are a chapter on 'Voiced and voiceless fricatives in English', identified as an English reworking of his Danish dissertation of the year 1891. —I have found nothing relevant in Wyld 1936 or in Chomsky-Halle 1968.
18. Note that I include the 2nd person singular pronoun. Apart from Jespersen, Jordan, and Dobson, the scholars quoted obviously did not do so in their statements, thereby neglecting the historical problem presented by this pronoun.
19. Or this is merely an ad hoc label of the kind discussed by Ohala 1985, pp. 234f.
20. I am grateful especially to Professor Gary Holland, who brought the Scandinavian data to my attention and guided me to the pertinent bibliography.

References

Annamalai 1968. E. Annamalai, 'Onomatopoeic resistance to sound change in Dravidian', in Bhadriraju Krishnamurti (ed.), Studies in Indian Linguistics (Professor M. B. Emeneau Saṣṭipūrti Volume) (Deccan College and Annamalai University, Centres of Advanced Study in Linguistics, 1968), pp. 15-9.

Anttila 1972. Raimo Anttila, An Introduction to Historical and Comparative Linguistics. New York: Macmillan Company.

Berger 1955. Hermann Berger, Zwei Probleme der mittelindischen Lautlehre. München: J. Kitzinger.

Bloomfield 1933. Leonard Bloomfield, Language. New York: Henry Holt and Co.

Bloomfield (Morton)-Newmark 1965. Morton W. Bloomfield and Leonard Newmark, A Linguistic Introduction to the History of English. New York: Alfred A. Knopf.

Bright 1972. William Bright, 'Hindi numerals', in M. Estellie Smith (ed.), Studies in Linguistics in Honor of George L. Trager (The Hague: Mouton, 1972), pp. 222-30.

Brough 1962. John Brough, The Gāndhārī Dharmapada. (London Oriental Series, 7.) London: Oxford University Press.

Burrow-Emeneau 1984. T. Burrow and M. B. Emeneau, A Dravidian Etymological Dictionary, 2nd ed. Oxford: Clarendon Press. Referred to as DEDR, by entry number.

Caillat 1980. Colette Caillat, 'Pronoms et adjectifs de similarité en moyen indo-aryen', in Indianisme et

bouddhisme: mélanges offerts à Mgr. Étienne Lamotte (Publications de l'Institut Orientaliste de Louvain, 23; Université Catholique de Louvain, 1980), pp. 33-40.
CDIAL = Turner 1966.
Chomsky-Halle 1968. Noam Chomsky and Morris Halle, The Sound Pattern of English. New York: Harper and Row.
Collinge 1978. N. E. Collinge, 'Exceptions, their nature and place—and the Neogrammarians', Transactions of the Philological Society 1978, pp. 61-86.
DEDR = Burrow-Emeneau 1984.
Diffloth 1976. Gérard Diffloth, 'Expressives in Semai', in Philip N. Jenner, Laurence C. Thompson, Stanley Starosta (edd.), Austroasiatic Studies (Oceanic Linguistics, special publication no. 13; University Press of Hawaii, 1976), pp. 249-64.
Dobson 1957. E. J. Dobson, English Pronunciation 1500-1700. 2 vols. with consecutive pagination. Oxford: Clarendon Press.
Emeneau 1951. M. B. Emeneau, Studies in Vietnamese (Annamese) Grammar. (UCPL 8.) Berkeley-Los Angeles: University of California Press.
—— 1969. ——, 'Onomatopoetics in the Indian linguistic area', Language 45.274-99 (reprinted in 1980, pp. 250-93).
—— 1978. ——, 'Towards an onomastics of South Asia', JAOS 98.113-30 (reprinted in 1988a, pp. 154-74).
—— 1980. ——, Language and Linguistic Area (ed. Anwar S. Dil). Stanford: Stanford University Press.
—— 1988a. ——, Sanskrit Studies ... selected papers (ed. B. A. van Nooten). Berkeley: Center for South and Southeast Asia Studies, University of California.
—— 1988b. ——, 'Proto-Dravidian *c- and its developments', JAOS 108.239-68.
Emeneau-Burrow 1962. M. B. Emeneau and T. Burrow, Dravidian Borrowings from Indo-Aryan. (UCPL 26.) Berkeley-Los Angeles: University of California Press.
Geiger 1916. Wilhelm Geiger, Pāli Literatur und Sprache. (Grundriss der Indo-Arischen Philologie und Altertumskunde I.7.) Strassburg: Karl J. Trübner.
Haugen 1976. Einar Haugen, The Scandinavian Languages: an Introduction to their History. London: Faber and Faber.
(von) Hinüber 1986. Oskar von Hinüber, Das ältere Mittelindisch im Überblick. (Sitzungsberichte der österreichischen Akademie der Wissenschaften, philos.-historische Kl. 467.) Wien, 1986.
Hock 1986. Hans Henrich Hock, Principles of Historical Linguistics. (Trends in Linguistics, Studies and Monographs 34.) Berlin: Mouton de Gruyter.
Hockett 1958. Charles F. Hockett, A Course in Modern Linguistics. New York: Macmillan Company.

Jespersen. Otto Jespersen, Selected Writings of Otto Jespersen. London: George Allen and Unwin, no date.
Jordan 1934. Richard Jordan, Handbuch der mittelenglischen Grammatik. (Germanische Bibliothek; 1. Abt.: Sammlung Germanischer Elementar- und Handbücher; 1. Reihe: Grammatiken, 13.) Heidelberg: Carl Winter.
King 1969. Robert D. King, Historical Linguistics and Generative Grammar. Englewood Cliffs, NJ: Prentice-Hall.
Kiparsky 1971. Paul Kiparsky, 'Historical linguistics', in William Orr Dingwall (ed.), A Survey of Linguistic Science (University of Maryland, 1971), pp. 576-649.
Krishnamurti 1961. Bhadriraju Krishnamurti, Telugu Verbal Bases: a Comparative and Descriptive Study. (UCPL 24.) Berkeley-Los Angeles: University of California Press.
Lehmann 1962. Winfred P. Lehmann, Linguistics: an Introduction. New York: Holt, Rinehart and Winston.
Mayrhofer 1956. Manfred Mayrhofer, Kurzgefasstes etymologisches Wörterbuch des Altindischen, Bd. 1. Heidelberg: Carl Winter.
Moore-Marckwardt 1966. Samuel Moore and Albert H. Marckwardt, Historical Outline of English Sounds and Inflections. Ann Arbor: George Wahr Publishing Co.
Ohala 1985. John J. Ohala, 'Around Flat', in V. A. Fromkin (ed.), Phonetic Linguistics: Essays in Honor of Peter Ladefoged (Orlando: Academic Press, 1985), pp. 223-41.
Pischel 1900. R. Pischel, Grammatik der Prakrit-Sprachen. (Grundriss der Indo-Arischen Philologie und Altertumskunde I.8.) Strassburg: Karl J. Trübner.
Postal 1968. Paul M. Postal, Aspects of Phonological Theory. New York: Harper and Row.
Robins 1978. R. H. Robins, 'The Neogrammarians and their nineteenth-century predecessors', Transactions of the Philological Society 1978, pp. 1-16.
Sweet 1888. Henry Sweet, A History of English Sounds. Oxford: Clarendon Press.
Tamil Lexicon. 6 vols. and supplement. Madras: University of Madras, 1924-39.
Thompson 1965. Laurence C. Thompson, A Vietnamese Grammar. Seattle: University of Washington Press.
Turner 1928. R. L. Turner, 'Romani les and Sanskrit tasya', BSO(A)S 5.43-51 (reprinted in 1975, pp. 310-18).
—— 1937. ——, 'Anticipation of normal sound-changes in Indo-aryan', Transactions of the Philological Society 1937, pp. 1-14 (reprinted in 1975, pp. 357-67).
---- 1966. ----, A Comparative Dictionary of the Indo-Aryan Languages. London: Oxford University Press. Referred to as CDIAL, by entry number.

—— 1975. ——, Collected Papers 1912-1973. London: Oxford University Press.
—— 1985. ——, A Comparative Dictionary of the Indo-Aryan Languages: Addenda and Corrigenda (ed. J. C. Wright). London: School of Oriental and African Studies, University of London.
Turner-Turner 1971. R. L. and D. R. Turner, A Comparative Dictionary of the Indo-Aryan Languages: Phonetic Analysis. London: Oxford University Press.
Whorf 1945. Benjamin Lee Whorf, 'Grammatical categories', Language 21.1-11.
—— 1956. ——, Language, Thought, and Reality: selected writings (ed. John B. Carroll). New York-London: John Wiley and Sons; Technology Press of MIT.
Woolner 1928. Alfred C. Woolner, Introduction to Prakrit, 2nd ed. Calcutta: Baptist Mission Press.
Wyld 1936. Henry C. Wyld, A History of Modern Colloquial English, 3rd ed. Oxford: Blackwell.