The development of "junk". Irregularization strategies of HAVE and SAY in the Germanic languages

DAMARIS NÜBLING

1. INTRODUCTION

Although it is a well-known fact that the most frequent verbs are the most irregular ones (if not suppletive), it is rarely asked how they became irregular. This article deals with the irregularization process of two originally regular (weak) verbs, HAVE and SAY in the Germanic languages, e.g. have, but has's and had'd (instead of regular *have's*/*have'd*) or say [sei], but says [sez] and said [sed] in English. Other verbs, such as do, go, stand, be, come, and so on, also tend to irregularizations again and again without any apparent reason. In contrast to HAVE and SAY these verbs have always been rather irregular, at least dating from their first written records.

Because little attention has been paid to this "regularity of irregularity", this fact has hardly been integrated in morphological theories (for such an integration see, however, Maiden 1991, Janda 1996, Lass 1990, Werner 1987a and b). This sort of irregularity cannot be subsumed under so-called "exaptation" (Lass 1990), that is, the functionalization of morphs which became nonfunctional, but rather under what Lass somewhat derogatively calls "junk" or "marginal garbage":

"Say a language has a grammatical distinction of some sort, coded by means of morphology. Then say this distinction is jettisoned, prior to the loss of the morphological material that codes it. This morphology is now, functionally speaking, junk; and there are three things that can in principle be done with it:

(i) it can be dumped entirely;

(ii) it can be kept as marginal garbage or nonfunctional nonexpressive residue (suppletion, "irregularity");

(iii) it can be kept, but instead of being relegated as in (ii), it can be used for something else, perhaps just as systematic.

[...] Option (iii) is linguistic exaptation." (Lass 1990: 81/82)

In this paper, we will deal with option (ii), but without sharing the characterization of nonfunctionality: firstly, this type of irregularity correlates too often with highest token frequency; secondly, it is produced too often using different strategies so that it cannot be considered undesired, nonfunctional, accidentally developed "waste from the past". Many morphological theories, such as naturalness theory (see Mayerthaler 1981, Wurzel 1984), consider irregularities to be the result of a "naturalness conflict", a (morphological) price for optimizations at the phonological level. Usually such phonological waste is regularized by analogical processes, but, in the meantime, it has been recognized that high token frequency can have preserving effects in particular if the phonological rule is no longer productive (cf., for example,
Verner's law, which was originally regular and whose last remainders can only be observed in some few verbs: New High German (NHG) ziehen – zog – gezogen (‘draw – drew – drawn’), English (Engl.) way – were).

Only the linguistic economy theory, which recognizes the benefit of such irregular relics, considers irregularity including suppletion to be increased formal distinctivity. Distinctivity permits a minimum expression without producing syncretism. Shortness of expression is extremely functional, especially under highest token frequency (see Zipf's law). Here, performance requirements such as short, simple sound chains move to the fore, while competence requirements such as rule-based, additively and transparently structured paradigms fade into the background. These competence requirements develop their effect at intermediate and lower frequencies and therefore concern most verbs, but not the peak level of frequency.

This article aims to demonstrate that irregularity not only develops (passively) by the preservation of nonfunctional relics (in the sense of Lass 1990), but can also be “created” actively by innovative processes. It focuses in particular on this second way of development. Considering that as yet no systematicatization and, correspondingly, typology of irregularization has been carried out, this article aims to trace the development of the irregularization of two important verbs, namely HAVE and SAY in ten Germanic languages, through the analysis of a relatively small sample. This is also intended to demonstrate that this phenomenon is valid crosslinguistically. Finally, we will consider the costs and benefits of morphological irregularity.

2. IRREGULARIZATION PROCESSES IN THE GERMANIC LANGUAGES: DIFFERENT PATHS TO “JUNK”

Irregularization processes become the most visible in cases where the verb was very regular before its irregularization (that is, it was not yet relatively differentiated, such as the strong verbs or the modal verbs). We will therefore analyze two originally (and, in some languages still) weak verbs, HAVE and SAY. Weak verbs are characterized by stable, uniform, agglutinatively connected morphs, such as NHG [lach]-[en] – [lach]-[t]-[en] – [ge]-[lach]-[t] ‘laugh – laughed – laughed.’ For this reason, in all Germanic languages, weak verbs represent the largest and most productive inflectional class into which many strong verbs transfer and to which all newly created and borrowed verbs are assigned. In contrast, strong verbs work with fusional (flectional) strategies and have – in particular in the modern Germanic languages – a clear affinity for irregular verbs since their tense ablaut (vowel change) is no longer predictable. The former seven Germanic ablaut classes have more or less split up to different degrees; today, New High German has about 30 different vowel alternations. Most of these alternation classes have only one member (see Augst 1975, Hempen 1988). Here, a high degree of irregularity has already been reached through natural processes. If this degree of irregularity correlates with a high token frequency, it tends to be preserved; otherwise, such “overdifferentiated” verbs transfer to more regular classes, mostly to the weak verbs.

In regard to diachronic depth, we will go back to the first written evidence. Since we must reckon with irregular developments at any time, it would be contradictory to use reconstructed forms which are always based on regular developments. In the case of the irregularizations, we will limit our analysis to only the most important strategies (for more details, see Nübling 2000). For reasons of space, we cannot list every individual paradigm (see, however, Tables 2 and 3, which contain the most important forms of HAVE and SAY).

2.1. HAVE in the Germanic languages

In all Germanic languages, a periphrastic perfect formed with the auxiliary HAVE (in some languages also with BE) developed about 1000 years ago. During its process of grammaticalization, HAVE became very frequent; today it shares the top region of the frequency scales with BE. As a consequence of this rapid increase of token frequency, HAVE has been extremely reduced and regularized in all languages.

In the New High German paradigm of haben, a longer stem hab[hab]-alternates in a unique way with a shorter stem ha-[ha]. In the present tense these "allo-stems" have "changing inflection" in the 2. and 3.sg.pres. of many strong verbs (NHG "Woehseflexion", i.e. the raising/leveling of e>1 in NHG gebe (1.sg.) vs. gibst/gibst (2/3.sg.) and the umlaut forms falle (1.sg.) vs. fällst/fällt (2/3.sg.)); that is, the finite form of NHG haben are distributed structurally analogous to the strong verb fällt with fälle vs. fällt/fällt in the singular and fällt, fallen, fallen in the plural. ich habe vs. du hast/sie hat (present singular) vs. wir haben, ihr habt, sie haben (present plural). The preterite and the conditional II have only short ha:-hatte 'had,' hatte 'would have.' In addition, the conditional II hätte is characterized by irregular (analogical) umlaut, since weak verbs were never affected by umlaut in the conditional. Here, analogy with the strong verbs must be assumed. These partially suppletive (since isolated) forms developed due to the Early New High German mixing of two paradigms which were still complete in Middle High German, namely of haben, which developed normally, and the reduced, irregularized hän. While Middle High German (MHG) haben preserved the old lexical meaning ‘have, possess,’ the formal reduction to hän correlates with the semantic reduction to tense forming ‘have’ as a perfect auxiliary. In Early New High German, this Middle High German paradigmatic splitting was abolished in favor of the creation of only one partially suppletive paradigm which now carries both meanings or functions again (see figure 1). Spoken German has further reductions and irregularizations: On the one hand, long [a] in the hab-forms is shortened to [a] (such as gehabt [go'hapt] ‘had’), while, on the other hand, bisyllabic haben contracted to monosyllabic ham [ham]. Such contractions are completely impossible in comparable verbs such as graben (→*gram). Contracted ham has become an unanalyzable portmanteau morph, which in turn increases the degree of allomorphy. At this point already, the duality of shortness and irregularity becomes obvious.
preterite stem hat-[hast]- diverges from this in that it has ablaut-like vowel change. This tense-specific vowel change emerged through the different but regular development of short a in the present versus long ā in the preterite. The shortness of the vowel in the present tense, however, is based on irregular reduction. Apart from these unique developments, the presence of a preterite is in itself remarkable since it only exists in the case of 10 to 20 verbs. Only in the imperative hieft! (sing.) and hieft (pl.) was the old stem final consonant preserved.

Dutch has relatively few irregular verbs. As a consequence, the irregularities of hebben are all the more important. The Dutch verb is characterized by syncretism in the second and third person of the singular – with exception of hebben ‘have’ (and zijn ‘be’): hij heeft [heeft] ‘you have’ vs. hij/zij heeft [heeft] ‘he/she has’; here, considerable stem allomorphs exist. The older form hieft resisted the usual analogical leveling to the second person singular with the result of a morphological overdifferentiation, which means that hebben distinguishes more different forms in the paradigm than usual. The preterite also diverges considerably from the weak pattern by having forms with a, that is, vowel-alternating forms (cf. the uniform sg. pret. had and the uniform pl. pret. hadden). This is no true ablaut, but rather the preserved case of so-called Rückumlaut (reversed umlaut) in Dutch. The umlaut forms in the present are the regular continuation of Germanic *habjan-. Since this į was missing in the preterite, no umlaut could develop. This vowel change was leveled out on the analogy of all the other verbs with Rückumlaut; Dutch does not even have a residual group of these verbs any longer, as in the case of NHG kennen – konnen, brennen – brannten, and so on. Dutch hebben thus behaves synchronically like a strong verb. This “strongness” is additionally supported by the (irregular) monosyllabicity of the sg. pret. had: if the form had developed according to sound laws, *hade should have resulted, as all weak preterites are bisyllabic (cf. hoorde ‘heard’). Following the pattern of the strong verbs (such as zong ‘sang’), however, Middle Dutch hadde, as the only weak verb, gave up the ending -e in the singular. In comparison, the bisyllabic preterite plural hadden corresponds to the general bisyllabic of both the strong and the weak verbs (such as zongen ‘sang’ (strong) or hoorden ‘heard’ (weak)). Finally, the whole preterite – just as in New High German, Swiss German and Letzehurgisch – is characterized by the (irregular) deletion of the stem final consonant s.

Swiss German (an Alemannic dialect) has reached a relatively high level of irregularization and reduction through completely different routes: Here only the Middle High German short-form paradigm hān has been continued. In the present, a sort of “changing inflection” arose, which in general does not hold for other modern Swiss German verbs: ich hā, du hesch, er hēt (present singular) vs. hānd (uniform present plural). The second and third person (hesch, hēt) contrast by primary umlaut from the rest of the paradigm, while the umlaut in the plural hānd is based on a later, morphologically conditioned umlauting. As in the case of New High German, the conditional has an irregular umlaut (hātt). Finally, the past participle is inflected like a strong verb (ghā ‘had’ < NHG gehān) – another adaptation to the strong verbs. In the whole Alemannic dialect area, the preterite has broken down and been eliminated.

Frisian has more reductions and, at the same time, more irregularizations than Dutch. According to Tiersma 1985 and Sjölin 1969, the following forms are the most important: ha [haz] or hawwe [‘hava] in the infinitive and in the uniform present plural, ha(w), ha-st, and ha-t in the present singular, the stem hie(-) [haz] in the preterite, and the past participle hān [haz]. Firstly, the stem final consonant disappeared almost completely – apart from some unique relics – which led to contraction and went in the direction of monosyllabicity. The preterite underwent particularly extraordinary developments: In Old Frisian, the verb abolished its preterite dental suffix and transferred to a small, irregular verb class with extremely high token frequencies; Old Frisian hāde/hedē → Modern Frisian hie. The concrete pattern for

Figure 1: MHG haben/hān and its continuations in Swiss German (Basel), New High German, and spoken German.
this strange process of analogy must have been *dwaan ‘do’ or *wêze ‘be’. Table 1 shows this small, interesting group and its preterites, which always contain the stem vowel ie:

<table>
<thead>
<tr>
<th>No.</th>
<th>Infinitive</th>
<th>Pret. sg.</th>
<th>Pret. pl.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>gean</td>
<td>gie</td>
<td>giene(n)</td>
</tr>
<tr>
<td>(2)</td>
<td>stean</td>
<td>stie</td>
<td>stiene(n)</td>
</tr>
<tr>
<td>(3)</td>
<td>dwaan</td>
<td>die</td>
<td>diene(n)</td>
</tr>
<tr>
<td>(4)</td>
<td>wêze</td>
<td>wie</td>
<td>wie(n)</td>
</tr>
<tr>
<td>(5)</td>
<td>ha</td>
<td>hie</td>
<td>hiene(n)</td>
</tr>
</tbody>
</table>

Table 1: Analogous processes in the preterite of Frisian verbs

The formerly weak verb ha(wwe) thus transferred entirely to the strong or, more exactly, to the irregular verbs of Table 1. This atypical analogical direction is supported by the strong past participle hán ‘had’ with the nasal suffix -n. English has split up its paradigm into full (independent) and enclitic (dependent) forms: on the one hand, have, has, and had (all pronounced with short [ə]) and, on the other hand, ‘ve, ‘s, and ‘d. Even the full forms show irregular reductions (deletion of stem final consonant -v in has, had, monophthongization and shortening of the stem vowel to [æ]). The orthography of <have> still implies the pronunciation of *[eɪ]; that is, have has divergent grapheme-phoneme correspondences (cf. the regularly developed behave). The bound, enclitic, syllabic minimal forms ‘ve [v], ‘s [s], and ‘d [d] are unique among the Germanic languages. They are used exclusively and accordingly frequently as auxiliaries. In this way, totally suppletive, nonsegmental minimal portmanteau morphs were created all at once. In English, the other three auxiliaries be, will, and shall are also affected by cliticization. As Krug (1994) points out, the clitic forms occur much more frequently than their full counterparts.

Starting with Danish, we will now deal with the North Germanic languages. Like all continental Scandinavian languages, Danish has very simplified verbal inflection (no person/number morphs). The finite forms of Danish have [hæ[ə]] are the following: pres. har [hør], pret. havde [hævd], sup. hæft [hæft]. The transcriptions show that the stem and consonant consonant has disappeared completely, except in the supine which — although being the only regularly developed form — has been isolated from the rest of the paradigm in this way.8

In Swedish, it is also the preservation of the old [ɪ] in the supine haft which uniquely makes the paradigm (as in Danish) more heterogeneous: hahtar [hɔː]/h[aː]r] (inf/pres. — hâde [hade]‘(pret. — haft [haft] (sup.). As in Danish, the loss of [ɪ] is irregular. In addition, a syntactic reduction (ellipsis) is extremely remarkable: In particular in subordinated clauses, the auxiliary ha can be totally deleted: han sa, att han (hade) varit sjuk (hade is usually omitted) (see Holmes 1994: 287).

In Norwegian (Bokmål and Nynorsk), too, have has been extremely reduced to ha/har — hadda — hatt. Here, however, fewer irregularities can be found than in Swedish. The supine hatt also lacks its stem-final consonant, which means that the former -v or -f was consistently abolished, but without any regular basis. Only the devoiced final consonant cluster -t (t[t]) in hatt diverges from the small, weak inflectional class into which ha has transferred (*hadda would be regular — cf. for example, bodd ‘lived’).

Faroese hava has developed a highly differentiated paradigm. Firstly, it shows a stable instability of the inflectional class, which means that, since Old Norse, this verb stands between two weak classes, the e class and the ja class. The infinitive and the uniform present plural, hava [hava], the first person singular present havi [havi], the conditional havi, the imperatives havi/havið, and the supine havi belong to the e class, while the second and third singular present follow the ja class: hevur [hø:vur] (*havir would be regular). The preterite has undergone extreme differentiation by splitting up the uniform singular and plural (which, however, is hidden by the orthography): hevði [heijj] (sing.) and hevðu [hø:d:v] (plural). Very complicated reductive and, at the same time, irregularizing phonological and analogical processes (which will not be explained in detail here) led to only the initial [h] remaining as the smallest common denominator. The preterite of siga ‘say’ and leggja ‘lay’ has experienced a similar development. As a result, a new small group has been established (see 2.2.).

As in the case of Faroese, Icelandic hafa shows “stable instability” by also standing between the ja class and the e class, but with different breaks. In Icelandic, the whole present tense (hef, hefur, hefur) belongs to the weak ja class. In addition, the conditional II contains — similar to German hätte — irregular i-umlaut: hefði instead of regular *hafði. Finally the supine hafa lacks the thematic vowel. As a result, Icelandic hafa has been isolated; apart from hafa there are no special reductions (see Table 2).

2.2. SAY in the Germanic languages

In contrast to have, say distinguishes itself by high token frequency without grammaticalization.9 Its increase in frequency can mainly be attributed to the decline of the strong Germanic verb *kwap- (OHG quedan, 5th ablaut class) and its replacement by the weak verb *sag-; interestingly, all Germanic languages underwent this change. Today, say belongs to the most frequently used verbs: NHG sagen is at Position 5 in the frequency of all verbs, Fris. zie at Position 6, Norwegian (Norw.) si at Position 11, and Icel. segja at Position 5.

In New High German, sagen — sagte — gesagt is a regular weak verb. German is thus the only Germanic language in which say has not undergone extraordinary developments. This was different in Middle High German and still is different in today’s Swiss German and other German dialects: Although OHG sagën (like habën) was actually a verb of the weak en class, it borrowed features from the weak
Table 2: The most important forms of HAVE in the Germanic languages

<table>
<thead>
<tr>
<th>Language</th>
<th>Infinitive</th>
<th>3rd sg./3rd pl.pres.</th>
<th>3rd sg./3rd pl.pret.</th>
<th>PP/sup.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1a) Alem.</td>
<td>hü [x]</td>
<td>hot(!)/hünd!</td>
<td>—</td>
<td>gäh</td>
</tr>
<tr>
<td>(1b) NHG</td>
<td>haben, (ham)</td>
<td>hart/haben (ham)</td>
<td>hatte/hatten!</td>
<td>gehabt</td>
</tr>
<tr>
<td>(2) Letz.</td>
<td>hun [u]</td>
<td>huet [vol]</td>
<td>hat/haten</td>
<td>gehat</td>
</tr>
<tr>
<td>(3) Dutch</td>
<td>hebben</td>
<td>heeft [e]/hebben [e]</td>
<td>had/hadden</td>
<td>gehad</td>
</tr>
<tr>
<td>(4) Frisian</td>
<td>havwe, ha</td>
<td>hat/havwe, ha</td>
<td>hie[lo]</td>
<td>hän [x]</td>
</tr>
<tr>
<td>(5) English</td>
<td>have, ‘well!’</td>
<td>‘it’ [i], from/’well’ [v], have!</td>
<td>‘if’, had!</td>
<td>had!</td>
</tr>
<tr>
<td>(6) Danish</td>
<td>have [ha][ which]</td>
<td>har</td>
<td>havde [‘hævde]</td>
<td>haft</td>
</tr>
<tr>
<td>(7) Swedish</td>
<td>ha</td>
<td>har</td>
<td>hade!</td>
<td>hatt</td>
</tr>
<tr>
<td>(8a) Bokmål</td>
<td>ha</td>
<td>har</td>
<td>hadde</td>
<td>hatt</td>
</tr>
<tr>
<td>(8b) Nynorsk</td>
<td>ha/have</td>
<td>har</td>
<td>hadde</td>
<td>hatt</td>
</tr>
<tr>
<td>(9) Faroese</td>
<td>hava</td>
<td>havor/hava</td>
<td>hovō [‘hovō]</td>
<td>havt</td>
</tr>
<tr>
<td>(10) Icel.</td>
<td>hafa</td>
<td>hafur (heftr)</td>
<td>hafa</td>
<td>hafa</td>
</tr>
</tbody>
</table>

Symbols: **Boldface**: Short forms, i.e. without stem-final consonant and (thus usually irregular)

<table>
<thead>
<tr>
<th>Underlining</th>
<th>Irregular form/segment (in regard to phonetics, orthography, and morpholology)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underlining</td>
<td>Irregular form/segment (in regard to phonetics, orthography, and morpholology)</td>
</tr>
</tbody>
</table>

**Boldface**: Short forms + irregular, whereby the irregularity does not + underlining: result from the reduction of the stem-final consonant.

The Development of “junk”. Irregularization strategies

The loss of the dental suffix -de-Ø (zeide → zei), however, is not explained; it lacks any regularity and only can be explained by analogy to the strong verbs which always are monosyllabic in the preterite singular. Because it also obviously contains a vowel change, zeugen belongs to the mixed verbs with strong preterite and weak past participle (gezeigt) in Hempen (1988: 27), whereby zeugen constitutes a separate subclass, which means that it is isolated from all other verbs. In the preterite plural zeißen, the dental suffix is still written, but it is already lost in spoken Dutch [zei] (see Haeseryn et al. 1997: 96).

Frisian sizze [stza] certainly is phonologically deviant, which, however, is not often treated in Frisian language histories. The regular Old Frisian palatalization of g- and k- before palatals must have arisen unusually in the middle of the word. The Old Frisian orthography <sed(s)za/sidza> tends to suggest the intermediate stage of affrication. The fronting of e > i, that is, the development from seggian to sizze, has not yet been explained (the same happened with leggian → lizzze ‘lay’). No other examples of e > i fronting has been documented for Frisian. Here, a so-called “over-palatalization” must have occurred, as in the case of other languages and verbs, such as give. Originating from the extremely palatal stem-final consonant g, an exceptionally regressive palatalization must have occurred from e > i. The present forms in the singular are sis [siz] ‘(I) say’, sei[s] [sait] ‘(you) say’, and sei[s] [sait] ‘(she) says’, of which the 2nd and 3rd person are based on (irregular) contractions. The uniform plural is homophonic to the infinitive sizze. As in the case of Dutch, the weak Frisian preterite has become strong through the irregular loss of its dental suffix and the monosyllabic singular form resulting from this: Old Frisian seide→Modern Frisian sei[st] ‘said (pret.)’. In opposition to Dutch, the Frisian past participle sein is also formed in accordance with the strong inflection (c.f. Old Frisian, where the past participle is still [e]seid ‘said’).

English grammars consider say to be an irregular verb and, more precisely, a verb with vowel alternation and a dental suffix in the preterite and past participle.
The reason for this is the monophthongization and shortening from [ei] to [e] also in the 3rd person singular present: say [sei], but s/he says [sez] and said [sed]. Engl. say shares especially the monophthongization in the 3rd person singular present only with very few, but characteristic verbs:

“Apart from the three primary verbs be, have, and do, the only verbs which have an irregular -s form are say /sə/ ~ says /sez/’, and derivatives of do [...]. In the -s form, say is irregular in pronunciation, but not in spelling. Gainsay, historically a derivative of say, may have a regular or an irregular pronunciation of the -s form: gainsays /sez/ or /sez/.” (Quirk et al. 1985: 99).

In this respect, says [sez] is more marked than said [sed].

Danish, like all Scandinavian languages, continues the Old Norse reverse umlaut verb sayga – sagði – sagt which, already at that time (as still in Modern Icelandic) contained “stable inflectional instability,” which means that it united forms of the weak ja and e classes (see below). On its way to Danish, sige [zi:o] has undergone three irregular developments: Firstly, it has become a short verb by consistently eliminating (apart from the supine sagt [sa:d] ‘said’ the stem-final consonant and by contracting to monosyllabic forms: siger [zi:er] (present) and sagede [ze:de] ‘preterite’ (the same holds for the two s-passive forms siges [zi:e] ‘is said’ and sagedes [ze:de] ‘was said’. Even the preterite suffix disappeared phonologically. Secondly, because the orthography continues the old extended form, there is a discrepancy between the written form and the pronunciation, which is extraordinary even for Danish (for example, see above). Thirdly, Danish (as in Norwegian) must have undergone a so-called “overpalatalization,” as is also assumed for Fris. szize: Under highest token frequency use, it can very often be observed that the interactivity between the sounds of a word, that is, contact phenomena, may appear which no longer follow sound laws and which do not occur in phonologically comparable words of lower token frequency. In this case, the palatal -g- regressive caused e > i fronting.10 Synchronically, Dan. sige should be regarded as a strong verb with a slightly suppletive supine.

Swedish has the following stem forms: sága [se:jà] – sa [so:] – sagt [sakt]: Here we immediately notice the short, monosyllabic preterite sa, which is in a constructionally counter-iconic relationship with the semantically unmarked, but formally marked present form såger [se:jè]. The situation for ångga ‘lay with’ with preterite la ‘laid’ is similar. The development was as follows: sågede > sade(he) > sae > sa (see Östman 1992). None of these processes of loss follow sound laws. With the loss of the dental suffix, this verb has become strong as well, as vocal change has also taken place between the present and the preterite. Finally, <sàga > [se:jà] contains an orthographic irregularity, as an application of the grapheme-phoneme correspondence rule, which is very clear and consistent in Swedish, would result in a pronunciation of *[sà:jà], that is, the (irregular) [g] > [j] palatalization is not represented graphically.

In Norwegian (Bokmål), si – sa – sagt is considered to be a strong verb with an irregular supine (Faarlund et al 1997: 186). Except for the bisyllabic present form sist [si:st], there are no more hints as to its past as a weak verb, not even in the orthography. Both the drastic reductions and the overpalatalization of e – i before palatals ([g]) are irregular.

In Faroese, siga [si:ga] has also undergone overpalatalization. In the present tense, some forms follow the ja class, while other follow the e class. Of particular notice is the number split in the preterite, which has led to similar forms as in the case of hav: sg. pret. segði [si:gi] vs. pl. pret. sagðu [se:gi]. Complicated regular and irregular processes of loss and partial analogies to various inflectional classes have led to this split, which is similar to the characteristic number ablaut split of today’s strong verbs. If the supine sagt and the passive form siger are integrated, one achieves a considerable inventory of differentiated stem allomorphs [si:]-, [si:], [sae:], [sak]-, [sik]- for a weak verb.

Since Old Norse, Icelandic segja has preserved the class instability between the ja verb and the e verb, but at a different mixture ratio than Faroese. Except for the missing theme vowel in sagt ‘said’, no special reductions have developed.

Table 3 provides an overview of the most important forms of SAY:

<table>
<thead>
<tr>
<th>Infinitive</th>
<th>3rd sg./3rd pl.pres.</th>
<th>3rd sg./3rd pl.pretr.</th>
<th>PPs/aux.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1a) Alem.</td>
<td>sige</td>
<td>sáit/sige</td>
<td></td>
</tr>
<tr>
<td>(1b) NHEG</td>
<td>sagen</td>
<td>sagen/sagen</td>
<td></td>
</tr>
<tr>
<td>(2) Letz.</td>
<td>soen, soen</td>
<td>soen/soen</td>
<td></td>
</tr>
<tr>
<td>(3) Dutch</td>
<td>zeggen</td>
<td>zeggen/zeggen</td>
<td></td>
</tr>
<tr>
<td>(4) Frisian</td>
<td>szize</td>
<td>szize/size</td>
<td></td>
</tr>
<tr>
<td>(5) English</td>
<td>say [sei]</td>
<td>say [sei]</td>
<td></td>
</tr>
<tr>
<td>(6) Danish</td>
<td>sige [si:]</td>
<td>siger [si:]</td>
<td></td>
</tr>
<tr>
<td>(7) Swedish</td>
<td>sága [se:jà]</td>
<td>såger [se:jè]</td>
<td></td>
</tr>
<tr>
<td>(8a) Bokmål</td>
<td>si</td>
<td>sier</td>
<td></td>
</tr>
<tr>
<td>(8b) Nynorsk</td>
<td>seier</td>
<td>seier</td>
<td></td>
</tr>
<tr>
<td>(9) Faroese</td>
<td>sige</td>
<td>sigur/siga</td>
<td></td>
</tr>
<tr>
<td>(10) Icel.</td>
<td>segja</td>
<td>segir/segja</td>
<td></td>
</tr>
</tbody>
</table>

Symbols: **Boldface:** Short form (and thus usually irregular)

"IRE": Shortness due to reduction of stem-final consonant

Underlining: Irregular form/segment (or, rather, phonetics, orthography and/or morphology)

**Boldface** + underlining: result from the reduction of the stem-final consonant.

Table 3: The most important forms of SAY in the Germanic languages
3. IRREGULARIZATION STRATEGIES AND OTHER PROCESSES AFFECTING HIGH-FREQUENCY VERBS

The described irregularization processes are limited not only to HAVE and SAY, but also affect other high-frequency verbs such as DO/MAKE, BE, GIVE, TAKE, BECOME, and GO. Grammaticalization seems to further these irregularization and reduction processes, but only insofar as they result in increased token frequency, which forms the actual drive behind these processes. Evidence can be found in SAY, which, strictly speaking, is not grammaticalized.

Basically, irregularization takes place on various linguistic levels:

- On the prosodic level (cliticization of English auxiliaries)
- On the phonological level (the many reductions, especially those affecting the stem-final consonants)
- On the orthographic level (deviations from the usual grapheme–phoneme correspondence rules)
- On the morphological level (see the various “atypical” analogies to (more) irregular verbs)
- On the syntactic level (see the ellipsis of the Swedish auxiliary ha in dependent clauses)

In the process, it is decisive that such irregularities arise not only “passively” through the accumulation of “junk” (as in Lass 1990), which was “overlooked” by analogy, but rather that they are also “actively” created: the latter is impressively substantiated by HAVE and SAY. Irregularities can thus be based on both conservative and innovative processes.

What are the morphological principles, then, which can be realized under very high token frequency? In summary, the following most important processes are let loose; in the process, we will examine the irregularization strategies of other high-frequency verbs (such as BE, GO, GIVE, and COME).11

Shortness of expression: Almost all changes in HAVE and SAY are partially connected to extreme reductions; that is, irregularity and shortness of expression usually appear together. Particularly subject to reduction is the stem-final consonant, such as NHG ha-st instead of regular *hab-st, Engl. has/ha-d instead of regular *hav-es/ hav-ed. Entire syllables often disappear, as in the case of the contraction of NHG haben → ham, Du. hadde → had ‘had’, Fris. hede → hie ‘had’, Swed. hava → ha, Engl. has → s, Old Norse (ON) segja → Norw. si, and so on.

Reductive phenomena can be ascertained on other levels as well: in the case of Swedish on the orthographic level (Swed. <hade> instead of regular *<hade>) and even on the syntactic level (the omission of finite ha in Swedish dependent clauses). It is very conspicuous that there are hardly any reductions which do not also cause irregularity: in most cases, the reductions spread only partially through the paradigms (and then in the most frequent categories). One of the rare examples of a systematic reduction (that is, without simultaneous irregularization) is the [ə] to [ə] reduction of Engl. have.

Irregularization: Almost every Germanic language12 has irregularized HAVE and SAY. In the process, principally two irregularization strategies have been used: (a) reductive irregularization and (b) nonreductive irregularization.

(a) Reductive irregularization: Through their only partial distribution, most reductions cause intraparadigmatic irregularization (whereby even the reductions themselves are also of irregular origin). In the process, stem-final consonant loss dominates. Honorous consonants are more apt to fall victim to this process than the more consonantal consonants, whereby this again is subject to token frequency: under extremely high token frequency, plosives can also disappear (such as Norw. la → late ‘let’). Within a paradigm, the stem-final consonant is more likely to disappear (especially in present singular) and/or then when complicated clusters would otherwise arise (see NHG ha be, but hast/ hat instead of *habst/ hab). Irregularization often also takes place due to the quantitative reduction of the stem vowel (cf. NHG hab- vs. ha-, Letz. hue- vs. hu-). Only seldom and under high token frequency is the initial sound reduced, which alienates and irregularizes the forms particularly strongly (see Engl. 3, ‘ve, d’). Psycholinguistic experiments confirm the particular prominence of the word-initial sound by means of word recognition tests (see Cutler et al. 1985, Fenk-Olofson 1989). Other reductions, such as those in orthography, also increase the irregularization balance (such as Swed. <hade> [ˈhade] with only one instead of the regular two <dd>). Of particular relevance is thus the observation that almost all morphological (analogical) irregularizations lead to increased shortness in expression; according to traditional thought, this is not the essence or the function of analogy. In the case of the “normal” so-called compensation or regularization analogy, longer expressions arise as usually more transparent structures are created; this is exemplified by the changeovers in class from the strong to the weak verbs (see NHG buk → back-er). On the other hand, the irregularizing analogies (“differentiating analogies”) documented in this article are always reductive (Old Frisian hade → Modern Frisian hie (after die ‘did’), MHG haben → hân (after gân ‘go’). In such cases, it is difficult to decide whether the analogy was possibly motivated by the shortness of the word.

(b) Nonreductive irregularization: This includes partial class changeovers, as has been practiced constantly by HAVE and SAY, especially in Faeroese and Icelandic. In the case of another verb, Swed. ge ‘give,’ former versions of the verb containing ə or i combined to a new and thus more strongly differentiated paradigm: ge/ger → gav → givit ‘give/gives – gave – given’. In Frisian, a unique analogy was carried out with the same verb, jaan ‘give’ to
a suppletive verb *slaan* ‘hit’ in the preterite: Old Fris. *jef* ‘gave’ developed to *joech* [juːχ] (in the entire preterite) under the influence of the suppletive preterite form *stoech* [sluːχ]. This has extremely differentiated the paradigm of *jaan*. In the case of strong suppletion through lexical blending, completely different paradigms are finally combined to a new one. This applies to *be* in all Germanic languages and to *go* in English: *go* – *went*. This means that, in the case of both irregularization types, there is a clear correlation between irregularity and shortness.

**Overdifferentiation:** In regard to the formal differentiation of grammatical categories, strong syncretism (homophones) in the high-frequency verbs should thus be expected due to the large number of reductions. Astonishingly, this is not the case at all – on the contrary: more and stronger differentiation is often made, particularly in the high-frequency area than anywhere else. Engl. *be*, therefore, has a triple differentiation of the sg. pres. *am/are/is* and of *was/were* in the usually uniform preterite. In Dutch, *hebben* (and *zijn* ‘be’) has an exceptional triple differentiation in present singular instead of the usual double differentiation: *ik heb, jij hebt, zij heeft*. Many New High German strong verbs and all weak verbs have syncretism between the 3rd person singular and the 2nd person plural present (such as *macht* ‘3rd sg./2nd pl. pres.’), but not for *haben* with *hat* [hat] vs. *habt* [ha:t].

In addition, formal differences can be more strongly contoured and thus more likely to be reanalysed or transported to the center of the verb than usual; see Engl. *says* [sez], where, except for the usual *-s* flexive, even the root is affected (through monophthongization and shortening).

**Intraparadigmatic positions of reduction and (over)differentiation:** Clear evidence for the token frequency argument (and against the mere grammaticalization argument) is provided by the position of the paradigm in which the reduction and differentiation takes place: in the process, these are usually the most frequent category combinations, such as the 3rd person singular present indicative active. In keeping with the trend, the following are more likely affected:

- The 3rd person (in comparison with the 1st and 2nd person)
- The singular (in comparison with the plural)
- The present (in comparison with the non-presents tenses)
- The indicative (in comparison with the conditional) and so on.

Even chronologically, it can be ascertained that changes appear first in the more frequent categories and later – if at all – in the less frequent categories. This categorial frequency is to be combined with the lexical frequency of the verb; that is, the preterite of *HAVE*, for example, is much more frequent than the present of *LAUGH*. The more frequent a certain verb is, the more possible it is for less frequent categories (such as conjunctive, preterite, 2nd person plural) to develop a deviant expression (see the following preterites: Engl. *was* vs. *were*, Far. *hevði* and *segði* [‘heuij/seji] (sg. pret.) vs. *hovdu* and *segdu* [‘hovdu’/se’odu] (pl. pret.), the New High German suppletive forms of *ist* (indicative) vs. *sei* (conjunctive)).

**“Boundary crossings”:** Under high frequency, the general tendency for various types of structures to become weaker or even disappear altogether has been observed again and again. This is reflected on many levels. On the prosodic level, for example, in the English auxiliary clitic forms of *have*, *be* and *shall/will* to ‘ve, ’s, ’m, ’ll, and so on, the word accent was abolished, and then the junctions disappeared; the result was the cliticization of these forms and their subordination under a new prosodic unit: *I have→I’ve [aiv], I am→I’m, I had/would/should→I’d, and so on.* The clitic forms must not necessarily – as in these cases – lose their syllable status; what is important here is only the loss of their status as a word and their word boundaries. On the phonological level, there is often interaction between the individual sounds that go beyond the normal dimensions, that is, the word-internal assimilation readiness increases. This includes the so-called “overpalatalization” of ON *segja* ‘say’ and *geva* ‘give’ to Norw. *si*, *gi*, Dan. *sige*, *give* (and borrowed from Danish as *give* into English), and Far. *siga*. In all cases, palatal g ([ɡ]) palatalized its vocalic environment beyond the usual dimensions by fronting the vowel e to i or by palatalizing it more strongly. So-called “overlabiovelarization” can be found in all languages in the case of the verb *COME*, where the original Germanic *kwem-* was assimilated to *kom-* and the initial consonant cluster was reduced. Bilabial [w] transferred the feature [+labial] to following e, which resulted in o. The only exception is the Dutch Pret. *kwam* ‘came’. Lower-frequency verbs with comparable phonetic conditions, such as OHG *quetan* ‘well up’ < Gmc *kwel-*, have retained the old vocal quality and initial consonance (NHG *quellen*).

An additional, interesting “oversimulation”, which occurs in various Germanic languages and which does not follow any sound laws, has appeared in the verb *COME*: before a following alveolar, the stem-final consonant *m* underwent an alveolarization to *n* in Alemannic, Letzbughish, North Friesian, and Icelandic (see Table 4): In Icelandic, this irregular assimilation is limited to the imperative singular form *<kómu*> [’kómu] ‘come’, which is hidden by the orthography.15 Not by chance, these irregular assimilations are carried out only in the present, even though the phonetic, but not the frequental prerequisites are the same in the preterite: Letz. *kéms* (Pres.) became *kéms*, but Letz. *koums* (Pret.) did not become *koums*. The paradigm is differentiated through such irregular and only partially effective processes (increased stem allomorphy). At the same time, through the creation of homorganic or even reduced clusters, articulation is facilitated.
Table 4: Assimilations in the case of COMB (Gmc *kwem-) in Alemannic, Letziburgish, and North Frisian

Under high token frequency, syntagmatic simplifications are carried out paradigmatically; that is, more importance is attached to articulatory economy than to uniform and transparent morphological structures. The respective cost-benefit analysis depends on the token frequency of the affected unit. Additional evidence for the phonological optimization at the cost of morphology and regularity is provided by the irregular assimilation and contraction of NHG haben→ham (in spoken German). This example clearly shows that such phonological Internal compression can lead to the dissolution of morphological structures: [hab]-[en]→[ham]. The highest internal compression at the total dissolution of morphological boundaries is carried out in suppletive forms (NHG bin ‘am’, ist ‘is’, sind ‘are’).

Additional boundary crossings or dissolutions are only hinted at. Even word-externally, increased external sandhi with syntactic neighbouring units may arise especially under high token frequency, without the verb itself criticzing (such as Alem. händheuer→hämäre ‘have we’). In regard to interparadigmatic aspects, the formation of strong suppletion due to paradigm mixing (such as go → went) represents a boundary crossing. On the semantic level, in the course of grammaticalization, so-called synsemantization takes place; that is, a grammaticalized realization through the combination of a finite auxiliary and a non-finite full verb.

4. THE FUNCTIONALITY OF “JUNK”: IRREGULARITY AS FORMAL DIFFERENTIATION WHICH PERMITS A MINIMUM EXPRESSION WITHOUT THE DANGER OF HOMONYMY

The common denominator of all the processes listed in Section 3 is that morphological and intraparadigmatic structures and rules are destroyed and, at the same time, the forms become shorter (or at least never longer). Irregularity and shortness also correlate to a high degree. Especially the morphologically destructive effects have led a number of morphological theories (such as the naturalness theory) to exclude the core area of these verbs from their general observations or to assign them a special status which isolates them from the rest of the “normal” verbs. This often happens to the verbs BE, HAVE, GO, and so on. In spoken German, only these verbs represent more than half of all verbs appearing in a text. For the further frequency ranking of German verbs, see Table 5:

<table>
<thead>
<tr>
<th>Verb</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>sein ‘be’</td>
<td>24.11%</td>
</tr>
<tr>
<td>haben ‘have’</td>
<td>22.72%</td>
</tr>
<tr>
<td>gehen ‘go’</td>
<td>4.77%</td>
</tr>
<tr>
<td>kommen ‘come’</td>
<td>3.78%</td>
</tr>
<tr>
<td>müssen ‘must’</td>
<td>3.24%</td>
</tr>
<tr>
<td>werden ‘become’</td>
<td>2.67%</td>
</tr>
<tr>
<td>machen ‘make’</td>
<td>2.58%</td>
</tr>
<tr>
<td>sagen ‘say’</td>
<td>2.26%</td>
</tr>
<tr>
<td>können ‘can’</td>
<td>2.01%</td>
</tr>
<tr>
<td>wissen ‘know’</td>
<td>1.21%</td>
</tr>
<tr>
<td>total</td>
<td>69.35%</td>
</tr>
</tbody>
</table>

Table 5: The frequency values of the 10 most frequent verbs in New High German (according to Ruoff 1990)

In order to be able to understand the function of irregularity, it is important to replace this negative term, which implies the absence of truly expectable regularity, by the positive term of distinctivity: irregularization always effects a differentiation of the paradigm, that is, the forms drift apart and become more and more dissimilar. This differentiation has the advantage of protecting the forms which become increasingly shorter under the effects of high token frequency from homophony (syncrhetism). Theoretically, with increasing word shortness, a merging of the forms should be expected, especially since the reductions – as shown – are usually at the end or in the middle of the word, where the most important categories are marked. The more strongly and further forwards the word is differentiated, however, the more strongly it can be reduced without the danger of homonymy. In its pure form, this ideal combination – minimum expression at maximum distinctivity – is realized by strong suppletive forms (cf. the English clitic forms of have, which are quantitatively minimal, but qualitatively maximally differentiated: ’d [d], ’ve [v], ’s [z]).

An entire scale of various coding techniques thus results – regulated according to token frequency – in accordance with the economy concept of Werner (1987, 1989) (see Figure 2 and the following quotation of Werner 1987b):
The Development of “junk”. Irregularization strategies

to the expanding class of weak verbs. The opposite case, which is much more seldom, is represented by HAVE and SAY. The observation of Vennemann (1993) explains these correlations as well (1993):

“The antithetic character of language changes is evident at other levels as well. For example, morphological iconicity is highly valued by speakers of languages; it is evidently related to the principle of uniform linguistic symbolization. But the iconic construction of words with complex meanings creates many long words, and length is a disadvantage on another parameter, especially for frequently used words. Thus we need not be surprised that the most frequent words of many languages are not iconic but suppletive and are, when lost, often not replaced with iconic constructs but with new suppletive forms. Paradigms [...] of words meaning ‘to be’ and ‘to go’ provide ample evidence for this in many languages.” (323)

The fact that the various coding techniques are conditioned by frequency and that every language has verbs of varying frequency explains why very few languages are typologically uniform. The described mixture principle is the most economical: “The aim of morphological change is a good mixture – not a uniform language type” (Werner 1987a). Only a balanced cost–benefit analysis which places both performance and competence needs in relation to token frequency leads to an adequate evaluation of the morphological relations and change and is able to integrate seemingly chaotic developments such as those presented in this article.

NOTES
1 I am very grateful to Martin Haspelmath for his useful comments on this paper.
2 For further criticism of these terms see Vincent (1993).
4 Already Old High German (OHG) haben – although originally a regular verb of the weak ein class – tends to different irregularizations. habend is the OHG verb with the most variants: Apart from various weak forms (OHG still has three weak inflectional classes), it also adopts features of the strong verbs, eventually develops contracted forms, and partially follows the small group of athematic verbs. In view of this excess supply of forms, OHG haben would have had many possibilities to adapt to a uniform inflection pattern, e.g. to the strong verbs as it is expected by Dishington (1980): “One must ask why the result was a mixed paradigm and not simply a strong one” (14). Instead, time and again, it tends to atypical, irregular, mixed paradigms.
5 The present tense of this short verb has leveled out analogically to the two stem verbs god ‘go’ and stáin ‘stand’, but not the past tense.
6 As already shown above for English, irregularities can also be found at the orthographic level. Especially Danish is well-known for its strong discrepancy between pronunciation and orthography. This discrepancy particularly affects frequent words, such as have [hæv]’. Even
in Danish, the correlation of <have> – >[hæv][1] is unique and highly unpredictable and must be learned as an exception. The same holds for the preterite havde – >[hæv][3][ø]. Only the supine haft [haft] has phonologically and graphically preserved the old stem-final consonant. Thus, only in written Danish has morphological uniformity been preserved for the most part.

1 In Swedish, the grapheme-phoneme correlation is more structured and closer to a one-to-one relationship than in Danish. Here, the only irregularity at the orthographic level concerns the unique writing of the preterite with only one instead of two <d> (cf. bodd ‘lived’ in regular writing). In Swedish, this orthographic irregularity is at the same time reductive while Danish preserved the writing of the older long forms.

2 The preterite <hadde> is (contrary to Swedish) correctly written.

3 Here, grammaticalization in a stricter sense is meant, namely as a qualitative process of the formation of grammatical categories, while Hopper/Thaugott 1995, for example, see the pure increase in frequency as a form of grammaticalization: “Frequency demonstrates a kind of generalization in use patterns” (103).

4 For a case of ‘overregularization’, see COME < Germanic *kwem- in all Germanic languages (except the preterite form kwam ‘came’ in Dutch), e.g. OHG queiman – NHG kommen, but OHG quelan > NHG quellen ‘swell’ (for further details and examples, e.g. Frisian jaan ‘give’, see Nübling (2000)).

5 For more details see Nübling (2000), where the reduction and irregularization strategies of ten verbs in ten Germanic languages are examined: HAVE, BECOME, GIVE, TAKE, COME, SAY, NE, DO, GO, STAND.

6 This finding also applies to the Romance languages. In what way all of these described processes are connected to the individual language or language type would be an interesting topic of research in a contrastive study.

7 On this point, see the Romance languages, which have mixed three verbs to form the cop paradigm (see Fr. vais – allons – irai).

8 There is usually syncretism between the 2nd and 3rd person singular present.

9 In many dialects such as Zuger German (chäisch, chüi) or the Cologne dialect (kuin), a segmental reduction has even appeared here. In Low German kümpe instead of *kumpe), the reverse, a progressive bilabialization of *r > p after m, has occurred (Münsterland; see Lindow 1998).

REFERENCES


The Development of “junk”. Irregularization strategies


Paradigm organization and lexical connections in the development of the Italian *passato remoto*

ELISABETTA MAGNI

1. INTRODUCTION

In modern Italian the verb “is a category whose complexity and intricacy is unmatched elsewhere in the grammar” (Maiden 1995: 122). On the whole the Italian verb system displays a historical tension between the tendency for one meaning to be represented by one morphological form (a tendency manifested through the mechanism of analogical leveling of allomorphy), and the amplification and spread of existing alternation patterns.

These phenomena can be observed in particular in the inflection of the *passato remoto* (remote preterite), where two different inflectional patterns, regular and irregular, exist. The second pattern shows stem alternations with highly idiosyncratic vocalic and consonantal allomorphy, and despite its complexity, it is not only preserved but also tends to become productive as time goes by.

The aim of this study is to investigate some unresolved problems connected with diachronic and synchronic aspects of this pattern of inflection. More specifically, on the one hand I examine mechanisms of morphological change of some innovative irregular preterites, and on the other hand, I discuss the strategies of alternating inflection and the possible reasons for its resistance to analogical leveling. I argue that the pathway along which the category of irregular preterites evolves and expands cannot be explained by assuming regular phonological developments but is crucially determined by morphologically and cognitively based generalizations. Following Bybee (1985), I also show that cognitive strategies are relevant in order to understand the way these structures of Italian verb morphology are produced, learned and changed by the speakers.

The proposal advanced here can contribute to the debate “connectionism vs. rules” (Dressler et al. 1997), where the issue under discussion is whether surface distinctions between regular and irregular forms have to be attributed to an underlying distinction in production mechanisms (single-system model vs. dualistic model, see Lazzeroni and Magni, forthcoming). In my opinion, the morphological systems of natural languages cannot be considered only in terms of a sharp dichotomy between absolute regularity and irregularity coinciding with the rule-rote bifurcation in lexical processing, and an alternative model of morphological processing is needed.

In fact, the regularity-irregularity split focuses on the two extremes of a continuum that, in synchrony and in diachrony, displays some intermediate areas of sub-regularity as well. Therefore I assume that, on the one hand, the regular aspects of inflection are covered by affixation that belongs to the system of combinatorial operations, and on the other hand the irregular domain of inflection is encoded in...