Analyzing strings into Morphemes

1 Finding a common morph—the Present Tense.

The first step is to find a common morph in a paradigm. The best way is to attempt to find the stem if that is possible. We will work with Russian. Consider the following paradigm, pre-arranged in person and then number: The paradigm is of the verb *nesti*, 'to carry (by foot).' Note: the phoneme /ö/ is always stressed:

Table 1

nesti	gloss
nisú	first person singular
nisöš	second person singular
nisöt	third person singular
nisöm	first person plural
nisöti	second person plural
nisút	third person plural

Look for common form of stem and add tentative morpheme there. After examining the paradigm, one should determine that the morph /nis/ forms the root/stem: [1]

Table 2

nesti	gloss
nis+ú	first person singular
nis+öš	second person singular
nis+öt	third person singular
nis+öm	first person plural
nis+öti	second person plural

[1] See the file on Root, Base, and Stem.

Table 2

nesti	gloss
nis+út	third person plural

Next consider the present tense conjugation of *vezti* ‘to carry (by conveyance).’

Table 3

vezti	gloss
vizú	first person singular
vizöš	second person singular
vizöt	third person singular
vizöm	first person plural
vizöti	second person plural
vizút	third person plural

As in Table 2 the morph /viz/ is determined, since it is common to all the forms:

Table 4

vezti	gloss
viz+ú	first person singular
viz+öš	second person singular
viz+öt	third person singular
viz+öm	first person plural
viz+öti	second person plural
viz+út	third person plural

Next, try find some more paradigms that are similar to the one above. For example, *vezti* ‘to lead:’

Table 5

nesti	gloss
vidú	first person singular
vidöš	second person singular
vidöt	third person singular
vidöm	first person plural
vidöti	second person plural
vidút	third person plural

Similarly, the morph /vid/ is determined:

Table 6

nesti	gloss
vid+ú	first person singular
vid+öš	second person singular
vid+öt	third person singular
vid+öm	first person plural
vid+öti	second person plural
vid+út	third person plural

We have now analyzed three stems: /nis/, /viz/, /vid/, each set has a single morph which may now consider an allomorph. But, there is more to come.

2 The Past Tense

Next we will look at the past tense paradigm of all three stems. In the past tense, number is not marked, but the opposition gender is marked. Gender is not marked in the plural:

Table 7

nesti	vezti	vesti	gloss
nös	vös	völ	past masculine singular
nislá	vizlá	vilá	past feminine singular
nisló	vizló	viló	past neuter singular
nislí	vizlí	vilí	past plural
carry	convey	lead	

Let's start with the second row, the past feminine singular. We find 'lá' in all three columns. We fin 'lö' in all three columns of the third row, the past neuter singular, and we find 'lí' in all three columns of the fourth row, the past plural. Note also that /l/ occurs in the forms except the past tense masculine of *nesti* and *vezti*. This wide distribution should lead us to hypothesize that /l/ is the default marker for the past tense. That leaves us to explain the absence of /l/ in those two forms. Assuming this, we will show the morphological analysis of Table 7 in Table 8:

Table 8[a]

nesti	vezti	vesti	gloss
nös	vös	vö+l	past masculine singular
nis+l+á	viz+l+á	vi+l+á	past feminine singular
nis+l+ó	viz+l+ó	vi+l+ ó	past neuter singular
nis+l+í	viz+l+í	vi+l+í	past plural
carry	convey	lead	

a. Below we will reanalyze the past masculine singular form as a nearly empty set.

Next we should observe that the hypothetical stem /nis/ occur in three rows of the four rows. In the first row, the past masculine singular, there is a phonological change in the tentative stem vowel. The past tense forms shares /n/ and /s/ with the tentative stem, the only change is in the vowel. It is rather reasonable to assume that there are two morphs: /nis/ and /nös/. Although there is not enough information to attempt to account for this alternation now, it is worth noting that it was said above that /ö/ occurs in stressed positions only. Here, the feature [-Stress] is in the same column as /ö/; the rule says to replace /ö/ with /i/ if it occurs with the feature [-Stress]. (5) and (6) are notational variants, because the rule will predict true results for either (5) or (6).[2] /s/ is consistent for the final consonant of the root/stem. We may hypothetically propose that the underlying form for *nesti* is {nös}.

Now let us consider the second column. We find a similar pattern, except we encounter a second morph /vös/. The voicing on the stem changes. We now have two allomorphs /viz/ and /vöz/. If we selected /z/ as the underlying form of the root/stem, we can easily predict that /z/ becomes voiceless at the end of a word: {vöz}

[2] For a proof of this, we need to go into deep logic.

In the third column we find that /v/ is consistent in the onset position, and we find the familiar alternation of /i/ and /ö/ in the peak position of the syllable: /vö/ and /vi/. In order to properly analyze the underlying form for *vesti*, we must include the present tense allomorph /vid/. Note that /vid/ occurs when the following segment is a vowel. Note that /d/ does not occur in the past tense, which is always marked with /l/. And we find the expected alternation of /i/ and /ö/. We are certain about the first two segments (the onset and peak of the morpheme): /vö…/. Is /d/ inserted epenthetically, or is it deleted? Arguments epenthetic insertion include:

(1) a. /d/ insertion is odd and rarely if ever found.

b. syllable-final consonantal deletion is not uncommon (the coda weakens).

c. Historically, Russian went through an open syllable phase: every syllable had to end in a vowel; this included diphthongs.

d. assume the underlying past tense feminine form: ##vöd+l+á##.

e. The old and perhaps the modern rule for syllabification is to start the syllable immediately after the vowel within a word:

##$ vö $ dlá $##.

f. There was a constraint in Indo-European that blocked an initial syllable cluster of a dental or alveolar stop plus /s/:

*$ {d,t}lV …$.

g. this constraint forced /d/ to be deleted. An epenthetic vowel could have been inserted, but the Slavic languages did not take this route.

h. The morphotactic structure of a root in nearly every Russian verb is monosyllabic and closed—i.e. it has a coda.

None of these arguments prove that there is an underlying /d/ in the root, but they constitute strong evidence for it plus deleting /d/ when it immediately precedes /l/ in syllable initial position. The three stem/roots for these three verbs are now:

(2) {nös, vöz, vöd).

What is the morpheme representing the past tense in the masculine singular word forms? /l/ occurs in /vö+l/, so there is no problem here. What about the past masculine forms /nös/ and /vös/? There are a number of linguistics who abhor 'Ø' as a morpheme. This used to include me. These days I am getting more and more interested in set theory. Set theory permits sets with small bits of information which cannot be pronounced. We will the set as '{f}', where 'f' stands for a feature. The feature in this text is either [+Past] (for past tense) or [-Fem, -Pl]. For reasons I cannot get into, '{f}' is not the same thing as '{Ø}'. Here the set has one member—'Ø'. When a form is deleted, I will assume the result is a nearly empty set '{f}'. Thus, the past tense has two allomorphs: {/l/}, and {[+Past]}. The morphological past tense masculine of the three verbs is now written as:

(3) nös+{[+Past]}, vöz+{[+Past]}, vö+l.[3]

In the past tense masculine, if /l/ follows an obstruent in word final position, /l/ must be deleted, has Russian has no other alternative: there is no syllabic /l/, and stops with a lateral release (t^l, d^l) are not permitted in syllable final position. The derivation of the 1st person singular and the masculine and feminine past tense word forms is shown in the following table:

Table 9

nös	vöz	vöd	root/stem
nös+ú	vöz+ú	vöd+ú	underlying Pres. 1st Pers. Sing.
nis+ú	viz+ú	vid+ú	vowel reduction = final form
nös+l	vöz+l	vöd+l	underlying Past Masc. Sing.
---	---	vö+l	delete /d/ before /l/
nös+{f}	vöz+{f}	---	delete final l after obstruent
---	vös+{f}	---	
nös	vös	völ	final phonemic form: '+' and '{f}' erased.

[3] Similarly, when a phonological segment is deleted, the result is an empty set in syllable position of the deleted segment. Below we show that /d/ is deleted before /l/. This would yield the past tense form of *vesti*: vi{}+l. I won't push this analysis here, however.

Table 9

nös	vöz	vöd	root/stem
nös+l+á	vöz+l+á	vöd+l+á	underlying Past Fem. Sing.
nis+l+á	viz+l+á	vid+l+á	vowel reduction
---	---	vi+l+á	delete /d/ before /l/

note that /d/ was deleted before /l/-Deletion. If the final /l/ were deleted before /d/-Deletion, the wrong from would be the result:

(4) vöd+l --> *vöd --> *vöt (word final devoicing.).

This is an example an ordered set of rules.

This leaves us with three tasks. Accounting for the morphophonemic changes that do occur, and how to analyze the past masculine singular., and finally, how to analyze the present tense endings. We will start with the obvious morphophonemic alternations.

On the one hand, we could list the distribution of the allomorphs in very general terms. Sometimes, that has to be done. But, recall, we have already noted that there are phonological conditioning rules, whether the be derived from an underlying form, or whether they are selected from a set based on phonological contexts. If a phonological context can be found, we should take advantage of it. One of the objectives linguists try to do is find an explanation for the things that happen. That isn't always possible, but if we can, we want to maintain our goal.

First, we will give the stress reduction rule for /ö/ in Russian as the following phonological rule of Russian (which we give for convenience):

(5) **Stress Reduction I (Russian)**

/ö/ --> /i/ / {_____, [-Stress]}.

Here the form enclosed in parentheses indicates a feature and the position where the feature occurs. This notation is used on computers and typewriters where vertical writing is difficult.[4] Often, you will find the feature and the feature written one over the other (vertical).

This is a notational variant of the one in (5). A notational variant refers to rule writing such that if one variant is true, then the other variant must also be true, and if one is false, the other must be false. This is a premise from logic.[5] Another notational variant is the following

(6) **Stress Reduction I (alternative, Russian)**

{/ö/, [-Stress]} -->/i/ / _____.

Here, only the relative set of features changes in the context given. If we combine the two ways of writing (5) (horizontally and vertically) and the two was of writing (6), we get four notating variants of writing a phonological rule. There are other ways, but we need not go any further into this here.

Let us return to Table 7. There are two allomorphs for the morpheme {NESTI}:[6] /nös/ and /nis/. In all forms except the past tense masculine, the wordforms for this paradigm are bisyllabic and the stress occurs on the final syllable. In the underlying morpheme hypothesis, suppose we write the underlying first person singular as ##{nös}+(ú)##.[7] The stress reduction rule is applied:

(7) a. ##{nös}+(ú)##

b. ##/nis/+/ú/##.

Each form is enclosed in curly brackets: '{' and '}'. The curly brackets represent an underlying morphophonemic form. Applying the stress reduction rule here creates a phonemic string, with the curly brackets replaced by the slants.

In the following section we will show that there are two allomorphs of the past tense morpheme: {/l/, {[+Past]}.

3 Simplified Set Theory Approach

---

[4] Chinese must be difficult to write on a typewriter as their system is traditionally vertical

[5] There are many notational variants used in various logical systems.

[6] We use caps to write the morpheme for the verb and use the infinitival form to represent the particular verb.

[7] '##' represents a word boundary. This form may be and is often shortened to ##nös+u##.

In the simplified set theory approach, there is no underlying forms but a set that contains at the moment two allomorphs: {/nis/ and /nös/}. In some sense, we have two sets in a string; the first set includes the verb stem allomorphs and the second one the first person singular (allo)morph:

(8) {/nis/ and /nös/}+{/ú/}.

Another fact about Russian. Only syllable in a word can be stressed.[8]

(9) **Basic Stress Pattern of Russian Word**
Only one syllable of a word may be stressed.

Now, we have to pick on the two allomorphs in the first set. In the second there is but one allomorph. We select that one first. Now we return to the first set. In view of Rule (9), we must select the unstressed allomorph. If we select the stressed allomorph, then the word will have two stressed syllables in violation of Rule (9). Therefore we get:

(10) ##/nis/+/ú/##.

Using set theory in this way, we don't have to make a claim that there is an underlying morphophonemic form which is never pronounced.

Now let us try to determine the allomorphs of the past tense. Note that /l/ is common to all but two of the past tense forms; /l/ does not occur in the past masculine singular of *nesti* and *vezti*: /nös/ and /vös/. There are two allomorphs for *vesti*: {viz, vös}. /s/ occurs in this paradigm in word-final position. With a larger corpus it can be shown that voiced obstruents never occur in word final position.Let us propose that voiced obstruents become voiceless in word final position. The proposed rule is:

(11) **Word Final Obstruent Devoicing**
[+Obstr] --> [-Voice] / _____ ##.

It turns out that this rule is found not only in Russian, but in several Slavic languages, German, and other languages in the world. As a matter of fact, partial obstruent devoicing

[8] Both parts of a compound word can be stressed, one stronger than the other.

occurs in English. For many speakers the devoicing process starts during the production phase of the final obstruent. For some speakers it starts a little sooner. However, the rule that lengthens a vowel before voiced segments in English remains. In this way the distinction between voiced and voiceless obstruents in word final position is maintained even though the vocalic lengthened bears the function now.

In terms of set theory we have two allomorphs for *vezti* mentioned above. Because of the constraint against voiced obstruents in word final position, the allomorph with the voiceless obstruent is selected when no suffixes follow:

(12) {/vöz/, /vös/}##.

Select the voiceless one:

(13) /vös/##.

The analysis of the past tense ending presents a bigger problem. In Table 7 as we have noted /l/ does not occur in the masculine singular forms of two verbs, though /l/ occurs in the past tense of *vesti*. Note that there are three allomorphs:

(14) {/vid/, /vö/, /vi/}.

Let us start with the first segment. Note that it is consistently /v/. We may posit /v/ in the underlying form. In the vowel position we find /ö/ and /i/, which is unstressed. Based on the alternations discussed above, we may posit /ö/ in the underlying form.

The fact that we are claiming that there are two past tense allomorphs: /l/ and [+Past] shows that the phonological approach to morphology is less than ideal. But we will have to let it stand for the moment.

4 The Past Masculine Singular Morpheme

We have not formally analyze the Past Tense Singular morpheme. Note that there is nothing following /l/ in /völ/, but the feminine, neuter, and plural suffixes are marked with /a/, /o/, /i/, respectively. Now that I have introduced the concept of set that contains one or more

features, we may analyze the past tense masculine as a set that contains only one feature, the feature marking the masculine gender, in all forms and that the empty set stands in contrast to /a, o, i/. The past tense masculine singular forms are rewritten as shown in (15), where {P} stands for [+Past] and {MS} for [-Fem, -Neuter, -Pl] (masculine singular):

(15) nös+{P}+{MS}, vöz+{P}+{MS}, vö+l+{MS}.

The first suffix after the stem/root represents tense and the second suffix gender and number.[9] Thus, it should be noted that there is but one allomorph for the masculine singular: {[Fem, -Neuter, -Pl]}.

5 The Present Tense Morpheme

Now I turn our attention to the present tense morpheme. Note that the endings for the 2nd person singular, third person singular, first person plural, and the second person plural all share the phoneme /ö/. Since there is a slot for the past tense, it is not unreasonable to look for one in the present tense. /ö/ is a reasonable candidate for this position. The four forms may be analyzed as:

Table 10

nesti	vezti	vesti	
nis+ö+š	viz+ö+š	vid+ö+š	2nd pers. sing.
nis+ö+t	viz+ö+t	vid+ö+t	3rd pers. sing.
nis+ö+m	viz+ö+m	vid+ö+m	1st pers. plural
nis+ö+ti	viz+ö+ti	vid+ö+ti	2nd pers. plural

6 The First Person Singular and the Third Person Plural Ending

[9] Historically, there was a vowel marking the masculine singular in the past tense. This vowel weakened and became phonetically null—but it left behind a feature in the set, a feature which cannot be pronounced.

The endings /ú/ and /út/ marking the first person singular and the third person plural, respective, are the same for all three verbs. Here, an empty set is not the best analysis for these endings. The approach which I favour is to assume that there is but one morpheme here and that this morpheme contains the features marking present tense, first person singular or the third person singular:

Table 11

nesti	vezti	vesti	
nis+ú	viz+ú	vid+ú	1st Person Singular
nis+út	viz+út	vid+út	3rd Person Plural

We will analyze the past tense and non-progressive participles of English in a similar fashion.[10]

A morpheme contains a bundle or set of grammatical features. Grammatical features may occur anywhere in the word or even in a clitic. We won't cover clitics until later. First, let us suppose that in the first morpheme following the verb in the first person singular and the third person plural forms contain the features [-Past], [-2nd], +[Pers], [-Pl]. This bundle of features is a set, and it is spelled out as the phoneme /ú/:

(16) {[-Past], [-2nd], [+Pers], [-Pl]} = /ú/.

A similar set (morpheme) exists for the third person plural ending /út/:

(17) {[Past], [+Pers], [-Pl]} = /út/.

Let us number the position for the stem as '1', and the suffix that immediately follows it as '2'. /ú/ and /út/ both occupy position 2.

[10] in set theory two sets overlap, some members of each set are shared by both sets. If both sets complete overlap, then the members of each set belong to both sets simultaneously. This is in effect one set. This is what is happening here.

Now let us suppose the feature bundles split into two morphemes or sets in the remaining forms of the present tense. We will assign tense to position '2a' and agreement to '2b'. Agreement includes those features which are subject to agreement rules with the subject of the sentence: person, number, and gender:

(18) Splitting in the Russian Verb
$[_2$ Tense, Agreement] --> $[_{2a}$ Tense] + $[_{2b}$ Agreement].

Note that all the past tense endings split this way into two morphemes via Rule (18): {nös+l+{}}. {nös+l+a}, and so forth. Let us now do a sample derivation of the second person singular form /nisöš/. First we start with the underlying form:

(19) {nös+$[_2$ [-Past], [+2nd, [+Pers], [-Pl]}

The splitting rule applies here creating two morphemes, the first for tense, the second for agreement:

(20) {nös+$[_{2a}$ [+2nd]] + $[_{2b}$ [+2nd] [+Pers], [-Pl]]}

The first morpheme is spelled out as /ö/ and the second morpheme as /š/.

This is one of the problems we run into when we try to define morphemes in terms of phonemes rather than as bundles or sets of grammatical features. Applying the splitting rules (18) and then the spell-out rules as indicated. We won't go into doing the above rules here as a class exercise. What we will do is set up an equivalence relation:

(21) **Correspondences for the Russian Verb**
{[-Past], [-2nd], [+Pers], [-Pl]} ↔ /ú/
{[Past], [+Pers], [-Pl]} ↔ /út/.
{$_{2b}$ [-Past], [+2nd, [+Pers], [-Pl]} ↔ /š/
{$_{2b}$ [-Past], [-Pers], [-Pl]} ↔ /t/
{$_{2b}$ [-Past], [-2nd], [+Pers], [-Pl]} ↔ /m/
{$_{2b}$ [-Past], [+2nd], [+Pers], [-Pl]} ↔ /ti/

The first correspondence that the first person singular is spelled out as /ú/, and the second as /út/.

These correspondences also work in reverse. If you find {/ú/}, then you know you have the first person singular form. Things get more complicated as more text comes in, but we needn't worry about it now. Correspondences are simpler than a rule format which I won't cover here at this time.

- Definitions
- Analysis and Rules 1
- Analysis and Rules 3
-

This page last updated 14 FE 2004

-