Analyzing ambiguity in orthographies

by Peter Unseth

Carole Unseth

© 2018 SIL International® All rights reserved.
Fair-Use Policy: Publications in SIL's Notes on Literacy series are intended for personal and educational use. You may make copies of these publications for instructional purposes free of charge (within fair-use guidelines) and without further permission. Republication or commercial use of SIL's Notes on Literacy is expressly prohibited without the written consent of the copyright holder.

Originally published as:


[Topics: ambiguity]

1. Introduction

It is common to speak of “ambiguity” in orthographies. We impressionistically quantify this, speaking of a “high degree” or “low degree” of ambiguity. We would like to define and differentiate two dimensions that contribute to the “degree of ambiguity” in orthographies. After discussing the two dimensions of ambiguity, we will also differentiate between two types of ambiguity. One of these two types will be further subdivided, though we admit that these may not be totally discrete categories.

This paper describes ways to analyze ambiguity, but does not directly address the problems created by ambiguity in orthography. There are greatly differing opinions on how much ambiguity an orthography can and should have, though it is unlikely there is one standard answer to all situations. We do not involve ourselves in this debate. Instead, our contribution is in providing concepts to allow for a more precise, informed discussion and judgment on such matters.

This paper treats reading and the recognition of words as being very closely linked to pronunciation and phonological processes. Fluent readers generally process written forms in different, more efficient, and less phonological ways, but we believe that this paper addresses the problems of orthographic ambiguity in a way that more closely reflects the problems and mental processes of new readers. The problems of orthographic ambiguity can also affect fluent readers (though they will be more likely to utilize context and redundancy). “The optimal orthography for a beginning reader is not the same as for a fluent reader” (Dawson 1989:1), but both can be hindered by ambiguity.


Page 1.
When literacy workers and linguists decide whether certain features will be marked in an orthography, such as tone, they usually try to measure whether the omission of a certain feature creates too many cases of ambiguity. A good knowledge of the language, together with proper testing, will help decide how many ambiguous words (or sentences) are created if the feature is omitted. If the omission of a certain feature does not create many potentially ambiguous words or sentences, the orthography is judged to have a “low degree of ambiguity.” Though meaning is sometimes ambiguous, in such cases, if the number of such ambiguous cases is low, readers can (supposedly) rely on context to disambiguate meaning. On the other hand, if the omission of the feature creates a large number of possibly ambiguous words or sentences, then this orthography is judged to have a “high degree of ambiguity.”

2. Dimensions of ambiguity

We measure ambiguity in two dimensions, what we call frequency and depth.

2.1. Frequency of ambiguity

Frequency of ambiguity is related to the number of times that an ambiguous form is found in written material. That is, “How often will the reader encounter this ambiguity?” If every vowel in the orthography is ambiguous as to length (and length is phonemic in the language), then the “frequency” of ambiguity is high.

As with phoneme frequency counts, the frequency of the occurrence of ambiguity may differ between studies of text and studies of the lexicon. For example, Amharic has phonemic gemination, but the syllable-based orthography does not show gemination. In the Amharic lexicon, there are relatively few pairs of words that differ only by consonant gemination. In text, however, the number of ambiguous forms (what we will later identify as “semantically ambiguous”) is much greater than in the lexicon. This is based on two factors, first, the relative frequency of the verbs all ‘it is’ and al ‘he said’. Also, the active and passive forms of verbs from a verb class labeled Group II (when based on the “contingent” stem) are ambiguous in the orthography because they differ only by gemination. In this case, the frequency of ambiguity in text is significantly higher than in the lexicon. Careful study of this problem in other languages may lead to similar or opposite conclusions.

2.2. Depth of ambiguity

The other dimension of ambiguity is “depth.” Depth of ambiguity is related to the number of different values a symbol or digraph (or even trigraph) can potentially have. For example, in a two-tone language, if tone is not marked, every vowel (or syllabic consonant) will have a tonal ambiguity whose depth is two, for example, bák and bāk. If, in addition to omitting tone, the orthography for this two-tone language also omitted vowel length, the depth of ambiguity for the orthographic form bak would be four, for example, bāk, bā̀ák, bā̀k, and bā̀àk. (If heterotonic vowel sequences were found in the language, the depth would be six, allowing bā̀àk and bā̀àk.)

Page 2.
As more phonemes are not symbolized in the orthography, the depth of phonological ambiguity increases greatly. Gudschinsky asked, “Is it possible that ambiguity increases in geometric proportion to the number of contrasts eliminated?” (Gudschinsky 1970:22). The answer to her question is “yes,” if we interpret her question as referring to the “depth of phonological ambiguity.” That is, in the above example, the loss of one contrast (tone) produced a depth of two, the loss of length added to the loss of tone produced a depth of four. The loss of a further phonemic contrast, for example, vowel nasalization, would produce a depth of phonological ambiguity of eight. In this example, the depth of ambiguity is calculated for each syllable, so polysyllabic words could have a very great depth of phonological ambiguity (though each syllable may have a lesser depth of ambiguity).

Frequency, then, is the number of times that an ambiguity is found. The frequency of an ambiguity may be quite high, and may vary between studies of lexicon and text. Depth is the number of possible values that an orthographic symbol (or combination of symbols) may have. The depth of an ambiguity will usually be a low number. The frequency of an ambiguity will almost certainly be a much larger number than the depth of an ambiguity.

3. Two types of ambiguity

Having introduced the two dimensions of ambiguity, we will now differentiate two types of ambiguity. The first, and most common type, is “phonological ambiguity.” The second, which derives from phonological ambiguity, is “semantic” ambiguity.

3.1. Phonological ambiguity

Phonological ambiguity is the result when a phonemically distinctive feature is not marked. When a phonemic feature is not marked, there will be written forms that have more than one possible pronunciation. When a written form has more than one possible pronunciation, this is phonological ambiguity. This does not mean that all of the potential pronunciations of a form have possible meanings, but it does mean that a reader must evaluate options and decide which possible phonological interpretation of a written form is correct. For example, in a two-tone language, when the orthography does not show tone, a written form, such as nibó, will have a depth (as defined above) of four possible phonological interpretations:

nibó nibò níbó níbò

Though only one of these may be a possible word in the language, the reader is still faced with phonological ambiguity when confronted with the form nibó. Because a phonemic feature of the language is not marked, the reader is forced to sort various possible interpretations and assign proper phonemic values to the ambiguous segments.

Though every example of a certain segment or sequence may be phonologically ambiguous, that does not necessarily mean that readers will always have to consider all possible phonological values for each example of the orthographic ambiguity. That is, though a form may have a depth of two or three possible values, the reader will not always consider each option. Often, one of the values of the depth of ambiguity (1991). Notes on Literacy, 65.
will be much more frequent, and thus be the usual default interpretation of that segment or sequence. For example, in English, the orthographic sequence *th* can be pronounced either voiced or voiceless. A native speaker of English will generally assume that it is voiceless, since this is the more common pronunciation (functors being a class of exceptions).

As a further example, in Gumuz, the implosive velar stop is very rare (merged totally with the egressive stop in some dialects). The current orthography does not distinguish between implosive and egressive velar stops. Though the symbol for this stop is technically ambiguous, with a depth of two (both egressive and implosive), readers will automatically assume the more common egressive interpretation, unless context forces them to reconsider this.

### 3.2. Semantic ambiguity

When a language has phonological ambiguity, it will almost inevitably also have semantic ambiguity. Semantic ambiguity is the result when more than one of the potential interpretations of a written form are possible words in the language. In the example above with *nibo*, let us assume that only one of the potential phonological interpretations of *nibo* was a possible word in the language. This represented only phonological ambiguity, because there was no other meaning possible from the written form. Now imagine that *nibô* means ‘tree’ and *nibò* means ‘egg’. In this case, the phonological ambiguity has given rise to semantic ambiguity, since there are now two potential meanings for the form *nibo*. All cases of semantic ambiguity arise from phonological ambiguity (except in the cases of homographic homonyms).

Not all cases of phonological ambiguity result in semantic ambiguity. However, since languages strive for efficiency and do not usually have “extra” phonemic distinctions, the total loss of a phonemic distinction in an orthography will almost inevitably lead to some cases of semantic ambiguity. It is possible to have a high frequency of phonological ambiguity and still have a low frequency of semantic ambiguity. If the phonological ambiguity is greatly restricted (for example, word-final high tone on short vowels is not distinguished from low tone if preceded by a glottal stop, or vowel nasalization is not marked orthographically on sentence initial vowels), then semantic ambiguity may not be found at all.

In addition to phonological ambiguity, when written words have plural possible meanings, we refer to this as “semantic” ambiguity. The more possible meanings written forms may have, the greater depth of semantic ambiguity a written form has. For example, in a Mazatec language, without any tone marking, the sequence *site* would have twelve possible meanings ([Pike 1948:23](#)), an unacceptably great depth of semantic ambiguity.

Semantic ambiguity can be further subdivided into two general categories, (though there are probably many examples that will not fit neatly into this dichotomy). We would like to speak of “lexical” and “grammatical” ambiguity. Generally, this relates to roots versus inflected forms.

### 3.2.a. Lexical ambiguity

(Pike 1948). *Notes on Literacy*, 65.
Lexical ambiguity means that the orthographic forms based on different roots are spelled the same. For example, if two nouns are spelled the same in an orthography, then this is lexical ambiguity. In lexical ambiguity, it is the specific lexical item that is unclear. This is not in any way limited to items which are the same part of speech. For example, in a Mixtec language, without tone marking the sequence, *naa* could mean ‘mother’, ‘I’, ‘losing’, and ‘will be lost’ (Pike 1948:6–7).

It may also be that these forms are inflected, but it is not the form of the particular inflection that is unclear, but the root. As an example, in Majang, the singular forms *tútúkán* ‘egg’ and *tíítuítúí* ‘tree stump’ differ by both vowel length and a final syllable, but in the plural, the difference is only vowel length. Therefore, an orthography that does not differentiate vowel length (or tone) will produce lexical ambiguity with the orthographic form *tutukak*, which would represent both *tútúkàk* ‘eggs’ and *tíítuítúíkàk* ‘tree stumps’. In these two forms, the reader will guess that the word is a plural noun, but the orthographic form alone does not make it immediately clear to the reader if the root is ‘egg’ or ‘tree stump’. This is a case of lexical ambiguity on inflected forms, even though the singular forms would be spelled with no lexical ambiguity.

### 3.2.b. Grammatical ambiguity

As opposed to lexical ambiguity, there is also grammatical ambiguity. Instead of two ambiguous forms from different roots, if the precise inflection of a word is ambiguous, then this is grammatical ambiguity. For example, if the inflected past tense and the future tense are ambiguous in an orthography, then this is grammatical ambiguity. For example, let us assume the following Nayi forms were to be written without showing tone: *hááy* on high tone means ‘ear’, but *hááy* on a low tone means ‘water’. Without tone being marked, this would be an example of lexical ambiguity. Contrast this with the potential grammatical ambiguity in the following Nayi examples: *nkeonu* ‘my dog’ and *nkeonu* ‘our dog’ (unmarked vowels have mid tones) (Aklilu 1990:3–4). In the first pair, it is two uninflected nouns that are ambiguous resulting in lexical ambiguity; in the second pair, it is two inflected nouns, resulting in grammatical ambiguity, since it is not clear which inflection is intended.

Grammatical ambiguity can also be illustrated from English, where the orthography has a very low frequency of semantic ambiguity. There are not many words where the same written form may have more than one meaning, and all involve grammatical changes (that is, all the examples we have found, so far). For example, the past tense of ‘read’ is spelled the same as some other forms (for example, infinitive, imperative, and so forth), but it is pronounced differently. Other examples involve pairs of nouns and verbs, which generally differ only by accent, such as ‘permit’, ‘compound’, ‘excuse,’ and so forth. In all these cases, the depth of semantic ambiguity is only two, since there are only two possible semantic interpretations.

If some grammatical category (case, voice, tense, person, number, negative, and so forth) is marked in the language by some phonemic feature that is not unambiguously symbolized in the orthography (tone, length, nasalization, and so forth) then a grammatical ambiguity results. This is potentially very serious, but the degree of frequency must also be considered. For example, if an orthography allows a potential grammatical ambiguity only between a third person plural double causative negative interrogative and first person inclusive plural pluperfect subjunctive, there would not be a frequent occurrence of these
ambiguous orthographic forms. However, if at least one of the two (or more) ambiguous forms is frequent in the language, then the frequency of the grammatical ambiguity may become serious. If more than one semantic value of the orthographic form is frequent in text, (that is, more that one of the possible semantic interpretations of an orthographic form are commonly used) then the semantic ambiguity is more likely to be serious, especially if the ambiguity is grammatical. For example, in Majang, the near past tense suffix is *ko* and the near future tense suffix is *koo* (tone differing, but dependent on the verb class). If the orthography shows neither tone nor vowel length, this semantic ambiguity will have a high frequency since these morphemes are used frequently. Though it has a depth of only two, both values of the depth are common so neither can be automatically chosen as the default.

It is not always possible (nor necessary) to distinguish lexical from grammatical ambiguity, but it can still be a convenient concept. The difficulty of trying to maintain this dichotomy is illustrated by the following possible orthographic form from Majang: *ŋadi* ṣ. In an orthography which does not show vowel length, this would be the orthographic representation for four possible phonological values, each of which has a semantic value:

*ŋadi* ṣ ‘she/he is angry’
*ŋadii* ṣ ‘we are angry’
*ŋaadi* ṣ ‘she/he believes’
*ŋaadii* ṣ ‘we believe’

Though we could contrast the pairs of these and identify the ambiguity as being “lexical” or “grammatical”, it is more useful to speak of this case as having a depth of “semantic” ambiguity of four, rather than trying to categorize the ambiguity between every pair of words.

### 3.2.c. Depth of semantic ambiguity

The depth of semantic ambiguity is more difficult to describe precisely than the depth of phonological ambiguity. For lexical ambiguity, the depth may differ from one orthographic form to another. That is, it may be possible that one orthographic form has a depth of three lexical interpretations, but another may have a depth of only two, and another form may have no lexical ambiguity. For example, in Gimira, the orthographic form *sam* (with no tone marked) can mean ‘cabbage’, ‘to be useless’, and ‘glow’ (noun). But there are other forms for which only two meanings are possible, and others for which there is only one possible semantic interpretation.

For grammatical ambiguity, it may be possible to describe the depth of ambiguity more precisely. For example, “under the proposed orthography, all genitive case nouns are ambiguous as to singular or plural number,” giving a semantic ambiguity depth of two. It will sometimes be the case that such descriptions of ambiguity can be stated with phonological conditions, for example, “all vowel-final noun roots, when marked for genitive case, are ambiguous as to singular or plural number,” or “all verbs of tone Class Two are ambiguous between active and passive forms.” For example, in Trique, without tone marks on the first syllable, there would be ambiguity as to tense on “all verbs … whenever all syllables do not lower to


Page 6.
Though the depth of semantic ambiguity may be small, semantic ambiguity may still have a high frequency. That is, even if there are not many written forms that have semantic ambiguity, if these are common in written form, then the frequency of the semantic ambiguity may force the adoption of some way to disambiguate such forms.

In cases where the depth and frequency of semantic ambiguity is low, the feature is often not generally marked in an orthography. When an orthographic form with semantic ambiguity is frequent, and both (all) values of the depth of ambiguity could easily be found in the same context, it may be necessary to create some way to orthographically disambiguate these forms. For example, in the Gimira language of southwest Ethiopia, those who prepared an orthography decided not to mark most words for tone, relying on context to disambiguate such sets as *sam*⁵ ‘cabbage’, *sam*⁶ ‘to be useless’, *sam*¹ ‘glow’ (noun). However, the third person singular pronouns for masculine and feminine differ only by tone, so these two words are orthographically marked to show the difference (Mary Breeze, personal conversation).

### 3.3. Problems arising from phonological and semantic ambiguity

Traditionally, frequency of semantic ambiguity is the measurement of ambiguity that has usually been discussed. As long as an orthography did not create too many semantically ambiguous forms, the orthography was generally felt to have a low degree of ambiguity and to be adequate. Where genuine semantic ambiguity did exist, it was hoped that context and collocations would solve the ambiguity. For example, if a verb and derived noun in a given language both had the same orthographic form, it was hoped that the position in the sentence would disambiguate the two forms. The phonological ambiguity faced by readers has not usually been considered as long as it did not result in a high frequency of semantic ambiguity.

Phonological ambiguity leads to semantic ambiguity when minimal pairs (triplets, and so forth) arise from an orthography. This is what is traditionally and usually meant by *ambiguity*. Conversely, if no minimal pairs were discovered, an orthography was assumed not to have any ambiguity. Following the terminology and concepts presented in this paper, such an orthography would be presumed to have no semantic ambiguity, but would still have phonological ambiguity. (However, it is doubtful that there would be absolutely no pairs with any semantic ambiguity. Who can be sure that they have thought of all possible orthographic minimal pairs in a language?)

But it is important to consider that even if there is little or no semantic ambiguity, there still may be phonological ambiguity. Phonological ambiguity can contribute significantly to difficulty in reading, especially for new readers. When readers must frequently consider potential options of what sounds and meanings are represented by written forms, they must perform many mental tasks, sorting out possible interpretations and choosing the best ones. The reader is forced to constantly choose between different options, leading to backtracking, a slow reading speed and to fatigue.
The “depth” and “frequency” of semantic ambiguity may be small, but the “frequency” of phonological ambiguity can still be great.

For example, let us imagine a two-tone language. If all combinations of tones are possible, then a trisyllabic word would have eight possible pronunciations:

\[
mópisú \quad mópisù \quad mòpisù \quad mòpisú
\]

\[
mòpisù \quad mòpisù \quad mòpisù \quad mòpisú
\]

In this hypothetical example, though there may be little or no semantic ambiguity, there is a high frequency of phonological ambiguity, every vowel. The depth of the phonological ambiguity is limited to a choice of only two tones, but it is very frequent, every vowel being ambiguous. Therefore, though the depth of phonological ambiguity may be small, the frequency of phonological ambiguity must also be considered in orthography preparation and evaluation. New readers, especially, who read by syllables at first would have difficulty.

There have been a number of “successful” orthographies which have had a fairly high frequency of phonological ambiguity. However, that does not mean that that frequency of phonological ambiguity can be ignored. A high frequency of phonological ambiguity may well be a significant factor in many less successful orthographies. Also, these successful orthographies might have been even more successful if they had had less ambiguity. (The question of measuring the success of an orthography is another matter!!)

For example, the Oromo orthography has been in use for about one hundred years, mostly in Christian churches. Though it is widely used, and many people can read it fluently, some find the orthography difficult. The main problem is phonological ambiguity. The Oromo orthography, using the Ethiopic syllabary, does not show two phonemic features: consonant gemination and vowel length (on three of the vowels). Using our terminology, the Oromo orthography has a high frequency of phonological ambiguity. The above features of vowel length and consonant gemination have a depth of two each. There is a high degree of phonological ambiguity and consequently, there are some cases of semantic ambiguity, as well.

This is illustrated below with an example written in the Oromo orthography based on the Ethiopic script.
Of the more than 30 possible pronunciations, only one is an actual word, *kerkeruu*. However, the reader is still faced with a high number of possible alternatives regarding the pronunciation of this orthographic form. While there is no semantic ambiguity in this example, the phonological ambiguity is high.

With amazing frequency, we find references in the literature to marking tone, or some other “exotic” phoneme, “only on minimal pairs.” Aside from the fact that we wonder if the analyst is aware of all possible minimal pairs, this practice still creates a high frequency of phonological ambiguity, even if it does not create a high frequency of semantic ambiguity. Even “if there are no minimal pairs, and yet the language is tonal, tone is a redundancy feature and still might need to be written for easy reading” (Wiesemann 1989:15).

The “amount” of ambiguity in an orthography is not just the frequency of semantic ambiguity. The total amount of ambiguity includes the depth and the frequency of ambiguity, both phonological and semantic.

The presence of some phonological ambiguity in any orthography is likely; with some scripts, it is inevitable. As those who plan orthographies are forced to leave certain phonemic features unmarked, a certain degree of ambiguity is inevitable. With each additional phonemic feature that is unmarked, the phonological ambiguity increases, and therefore the likelihood of semantic ambiguity also increases. By considering the concepts of “depth” and “frequency” of both “semantic” and “phonological” ambiguity, it should be possible to make these decisions in a more informed, more thoughtful way.

### 4. Quantifying ambiguity

It might be appealing (to some) to devise a way to mathematically quantify ambiguity in orthographies. Then a certain standard could be established, for example, any orthography that has a score above $X$ is too ambiguous and must be improved. It would then be a mathematically straightforward task to take a piece of text, prepare it in the various proposed orthographic forms, then to calculate all of the various ambiguities and total them up. Gordon tells of an experiment where such procedures were actually tried, but some parts of the test ended up being “arbitrary” and some calculations became “prohibitively difficult” (Gordon 1986:76).

We do not think such a procedure is possible or useful. Such a measurement would appear to give the precise mathematical measurement of ambiguity in an orthography, giving the illusion of precision. This could not give a true measurement of the real ambiguity for a *native speaker* reading the text. Especially for fluent readers, many potential ambiguities (especially phonological ambiguities) will not even be considered due to redundancies in the language and context. For example, if two nouns are written the same, but one means ‘weed’ and the other means ‘eclipse’, a text about caring for bean fields will automatically provide the context to prompt the reader to instantly choose the pronunciation for ‘weed’, (1991). Notes on Literacy, 65.
never even considering the other option. Also, if verbs and derived nouns are differentiated only by tone, in a verb-initial language, the (fluent) reader would quite automatically choose the verbal tone pattern for the first word in a sentence, and the nominal tone pattern for a word that followed a verb. Though some things may be theoretically ambiguous to a linguist, a native speaker reading them may not sense the ambiguity.

As another example of native speaker intuition compensating for an orthographic ambiguity, let us briefly present gemination in Amharic verbs. Gemination within verbs is closely related to verb classes and inflections. For most verbs, the gemination is redundant, the affixes and root-internal vowel inflections can unambiguously differentiate the particular verb form. Therefore, a native speaker who is a fluent reader can usually deduce the consonant gemination quite quickly, even automatically, even when reading aloud. This suggests that the native speaker who is a fluent reader may not be greatly hindered by a phonological ambiguity which can be resolved by other redundant features of the morphology.

For these kinds of reasons, it does not seem possible to produce a useful, accurate way to mathematically quantify ambiguity based solely on analysis of written texts.

Though we have just discussed ways in which context may help a reader to correctly interpret an ambiguous orthographic form, we believe that readers, especially new readers, will not and cannot take as many clues from context as some orthography planners may hope they will. There appears to be no substitute for the careful testing of ambiguity in orthography.

5. Seriousness of ambiguity

Not all cases of ambiguity are equally serious. For instance, as mentioned above, in Gumuz, the implosive velar stop is very rare, in some dialects merged totally with the egressive voiced stop. Therefore, the orthography for Gumuz does not differentiate between implosive and egressive velar stops. This does theoretically create phonological ambiguity for every use of the voiced velar stop symbol, but in practical terms, the readers will normally assume the egressive phoneme and not be hampered by the possibility of the rare implosive stop. As far as we know, there are no examples of semantic ambiguity that arise from this. Therefore, this case of phonological ambiguity is of little consequence.

On the other hand, some cases of orthographic ambiguity would be catastrophic. For example, if tone were not marked in the Mazatec example given above, the frequency and depth of ambiguity, not merely phonological but semantic, would make it impossible to read it. There is a tendency to “simplify” an orthography by not marking tone (other features as well, but most often tone) if there are not many minimal pairs. “Though tone may not provide many words in a language differing by pitch alone, it may nevertheless play an important part in the language. It is a mistake to ignore the tonemes of a language just because few words depend on them entirely to distinguish meanings” (Pike 1948:6–7).

As linguists try to gauge the seriousness of orthographic ambiguity, they must be wary, for “the determination of functional load depends upon what clues the speakers of a language actually use to distinguish pairs of similar words” (Gudschinsky 1973:120). What may appear to be a minor ambiguity (to the linguist) may in fact be a major ambiguity to the native speaker/reader, and vice versa. Though the
omission of a phoneme may create few minimal pairs, this phoneme may be one that native speakers focus on to a high degree.

Grammatical ambiguity would appear to be more serious than lexical ambiguity, especially if it would appear more frequently in text and could not be resolved as easily by referring to context (see the examples above of the Majang near past and near future suffixes).

Lexical ambiguity would probably be more serious if the potential meanings of an orthographic form were the same part of speech, for example, nouns. Presumably, if a verb and a derived noun were spelled the same, their position in a sentence would help disambiguate these homographic forms. However, new readers will not utilize the contextual clues as effectively as fluent readers.

The position in the sentence can be crucial for grammatical ambiguities. If it occurs early in the sentence, it is often more crucial; if it occurs later in the sentence, then the initial parts of the sentence may provide enough context to help the reader to disambiguate (Gudschinsky 1973:121–122). For example, if the (unmarked) tone on a sentence initial morpheme differentiated imperatives and interrogatives, then the rest of the sentence would be understood (at least in the first reading pass) according to the reader’s initial interpretation of the tone on the sentence-initial morpheme.

This situation is found in Murle (a language of the Sudan-Ethiopia border), where tone has been found to carry so little functional load that it is not marked in the orthography, except on the sequence ma. Jon Arensen explained it to me as follows: “This word comes at the beginning of a clause and depending on tone and fortis can mean ‘and’, ‘don’t’, or ‘if’. The entire intonation of the following clause depends on getting the first ma pronounced correctly. In most situations the context helps the reader understand how to pronounce a given word, but since ma almost always precedes the clause, the context is of little help until the clause is read several times. Even then it can be totally misunderstood. Therefore I have chosen to write these three words with diacritics and these words are taught as separate sight words in the Murle primer” (Jon Arensen, private communication, 1990). He gives the following examples, in the Murle orthography.

Má a ṣoli

‘Don’t be afraid!’

Mà a ṣoli

‘And he fears.’

Ma a ṣoli, ngaan aviir.

‘If he fears, he will run.’

6. Summary

We have explained how we distinguish ambiguity in two dimensions, frequency and depth. Frequency refers to the number of times that a particular ambiguity occurs, whether in studies of text or lexicon. Depth is the number of values or interpretations an orthographic form may have.

We have also distinguished two types of ambiguity, phonological and semantic. Phonological ambiguity means that a reader can pronounce a written form in more than one way, regardless of whether these alternatives have any meaning in the language. Semantic ambiguity means that of the possible interpretations of the orthographic form, more than one is a meaningful word in the language.

By presenting the concept of phonological ambiguity apart from semantic ambiguity, we have tried to show that an orthography can still present frequent ambiguities to a reader, even if little or no semantic ambiguity exists. Orthography planners cannot ignore the potential seriousness of this problem, especially for new readers.

We have divided semantic ambiguity into two subtypes, lexical and grammatical, though admitting that this is not always a discrete dichotomy. Lexical ambiguity is the result when an orthographic form allows for the interpretation of different possible words from different roots. Grammatical ambiguity is the result when an orthographic form is ambiguous as to its inflection.

The chief problem of ambiguity in orthography is that it may force a reader to reread a passage or it may actually lead to incorrect understanding of a passage. “In general, an orthography that forces a person to read something more than once in order to understand it is a poor one” (Gudschinsky 1973:127).

We have also argued that it is not possible to precisely quantify the ambiguity that a native speaker will face in an orthography, especially as the person becomes a fluent reader. There is no substitute for careful testing of an orthography, but we hope that the concepts presented here will help orthography planners to devise more useful tests.

References


Citations


View document See also Dawson, J. 1989 Bibliography (Literacy) Citations


Page 12.

Citations


Citations


Citations


Citations


Citations


Page 13.