2. MORPHOLOGICAL ANALYSIS: SEGMENTATION AND CLASSIFICATION

General Morphological analysis was developed by Fritz Zwicky

First step. The problem to be solved must be very concisely formulated.
Second step. All of the parameters that might be of importance for the solution of the given problem must be localized and analyzed.
Third step. The morphological box or multidimensional matrix, which contains all of the potential solutions of the given problem, is constructed.
Fourth step. All of the solutions contained in the morphological box are closely scrutinized and evaluated with respect to the purposes that are to be achieved.
Fifth step. The optimally suitable solutions are ... selected and are practically applied, provided the necessary means are available. This reduction to practice requires in general a supplemental morphological study.

Fritz Zwicky (1898–1974) was a Swiss astrophysicist (galaxy Zwicky 18 is named after him) and author of Morphologische Forschung (Winterthur, 1959) and Entdecken, Erfinden, Forschen im morphologischen Weltbild (München, 1966). His distant cousin, Arnold M[elchior] Zwicky, on the other hand, is an American linguist, and Arnold’s “How to describe inflection” (in the Proceedings of the 11th Annual Meeting of the Berkeley Linguistics Society, 1985) is even closer to our specific concerns here than Fritz’s all-purpose Morphological Box. Remember reading this paper as background to Chapter 7!
2.1. SEGMENTATION

Morphological analysis means, among several other things which will occupy us later:

- to divide up (or segment) whole words into their (smaller and smallest) constituent parts, with these parts themselves having meaning.

Dividing up words into syllables or sounds would be doing phonological analysis. Morphological and phonological word parts may coincide – and if they do so, this would be an interesting finding, especially if such coincidences aren’t random and coincidental.

Dividing up sentences into clauses, clauses into phrases, phrases into words is syntactic analysis, and we are here taking for granted that this has already been done and word units are ready for us to investigate.

These meaningful parts into which words can be segmented are called MORPHS. MORPHEMES are something else, something more abstract, as will be explained shortly. (Remember PHONEMES and PHONES/SOUNDS.)
For languages that we know well such segmenting can be done **intuitively**: Our linguistic intuitions (“Sprachgefühl”) tell us whether words are simple or complex, and if they are complex, which parts they consist of.

Do they?

Here are some English sentences, first with their sound form given in phonetic transcription and with their meanings provided through translations into German (some of them also through pictures):

meanings, in pictorial form
How is the sound stream, given in phonetic transcription, divided up (into 3 parts each)?

How are the German translations divided up (into 6 parts each)?
Now in standard **English orthography**, on the left-hand side (next slide), these sentences are divided up into “words”, separated by blank spaces.

What the correct word separation is would of course have to be demonstrated through convincing analysis: but for once let’s trust the orthography. Here we are only concerned with the **internal** grammar of words and largely take the grammar of phrases, clauses, sentences for granted.

In the **phonetic transcription** above there were fewer blanks than in orthography: the units of pronunciation separated by blanks here are defined (among other distinctions) through having one **main stress** each (= phonological words).

On the right side, we have added more blanks, to indicate those meaningful parts of words which our intuitions tell us to assume.
The cat sits on a mat.

The dog pees on a mat.

The cats sit on a mat.

The dogs pee on a mat.

The cat sat on a mattress.

The dog peed on a mat.

some complication here:
no simple segmentation into two successive parts!
Question:
Is the supposedly meaningful part \( s \) (this is how it is spelled; soundwise we have \([s]\) and \([z]\)) always the same? Or are there an \( s_1 \) and an \( s_2 \) (meaningwise)?
More questions later.
But: What are our intuitions based on?

Why don’t we, intuitively, segment these words too?

f at
c at
s at
m at

Why might we be tempted, intuitively, to segment further English words as follows?

fl are    gl are    st are    bl are
fl ow     gl ow    bl ow
fl ick    st ick
fl itter   gl itter    st utter
fl immer   gl immer
fl oat     gl oat    bl oat
Question:
And what about segmenting -eme in morpheme and phoneme? Any other such -eme words that would confirm such a segmentation?

Der Vollholz-Sessel “Morph” von Zeitraum ist gerade auf den Markt gekommen und hat bereits den “Interior Innovation Award 2011” gewonnen.
As linguists we want to understand what speakers-hearers “intuitively” know or feel about language(s). We take intuitions (introspective judgements) seriously and investigate them methodically, just as we investigate utterances themselves (language-in-use), in order to gain insight into the mental lexicon-and-grammar: the system and subsystems of forms and form-systems and of rules for and constraints on constructions that enable people to speak and understand a language – any language they are exposed to (when this exposure occurs at the right time of a person’s life).
How do we find out how such intuitions about words and their parts emerge?

Let’s create a situation where we can’t have any such intuitions, and let’s study ourselves gaining morphological intuitions/knowledge.

Let’s therefore look at (better would be: listen to) a language about which we know/feel little or nothing – except that it is a human language, a language humans are able to learn and speak and understand, with a morphology that it is humanly possible to master (a morphology that has been shaped by a possibly long history of linguistic developments, acquired and perhaps modified by one generation of language acquirers after another).
**Technical note:**

Hyphens (-) are used in morphological analysis to indicate boundaries between meaningful word parts. Sometimes also pluses (+).

Sometimes several boundary markers are used to distinguish different kinds (or strengths) of word-internal boundaries.

CLITICS (to be explained later) are usually separated off by the equal sign (e.g., English fast, informal speech *I=m spos=ta meet=ɔm* for *I am supposed to meet them*).

Labels for grammatical categories are often given in small capitals, often in abbreviation.

More details later – and one ought to be very careful about such matters of presentation.

In anticipation, study these instructions for (what is known as) interlinear morphemic glossing:

- Leipzig Glossing Rules
Say it in Kekché

‘you (SG) are going’ ____________________________
‘we will go’ _________________________________
‘you (PL) are going’ __________________________
‘we are ill’ _____________________________
‘you (SG) are a man’ __________________________
‘he is a man’ _________________________________
‘we are women’ _____________________________
‘you (PL) are women’ __________________________

That is: Express those meanings in Kekché which on the left hand side are expressed in English, enriched by some grammatical terminology.
Kekchi (also spelled Q’eqchi’) is a Mayan language spoken in Guatemala, Belize, and El Salvador.
15 minutes
to solve this puzzle
and to reflect on what exactly you’ve done solving it

Two flavors in one package - each changes into a new flavor because of "flavor changing beads." Well, it doesn't quite work that way.
You don’t know any Kekchí? Here are a few scraps of Kekchí and what they mean: Will this help?

‘I will go’ in Kekchí is \textit{tinbeq}

‘you (SG) will go’ \textit{tatbeq}

‘I am going’ \textit{ninbeq}

‘I am ill’ \textit{yašin}

‘he is ill’ \textit{yaš}

‘you (SG) are ill’ \textit{yašat}

‘you (PL) are ill’ \textit{yašeš}

‘I am a man’ \textit{gwinkin}

‘we are men’ \textit{gwinko}

‘she is a woman’ \textit{išk}
As you will have noticed from the way the examples are spelled (alphabetic script, but no internal blank spaces), Kekchí is one of those languages where single words suffice to express whole propositions (whole thoughts/Gedanken, if you prefer: something that can be true or false) – for whose expression languages like English need entire sentences (if relatively short ones).

Such comprehensive expressions, not showing an obvious division of sentences into words, could be called “sentence-words”.

In the respects relevant for us, though perhaps not in all others, they are truly words: and their analysis therefore is the responsibility of morphology.
PS: Not that this is necessary for present purposes, but if you’d like to know more about Kekchí, or any other language about which you know little or nothing, go google. If you’re really serious, however, you’ll want to consult the scholarly literature – grammars, dictionaries, monographs, and articles such as these, for the language at issue:


Translation:
Everyone is entitled to all the rights and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status. Furthermore, no distinction shall be made on the basis of the political, jurisdictional or international status of the country or territory to which a person belongs, whether it be independent, trust, non-self-governing or under any other limitation of sovereignty.

Source: http://www.language-museum.com/encyclopedia/k/kekchi
Manuscript in Kekchi, Quiché, Latin, and Spanish. Guatemala, c.1544-1570.
http://www.library.upenn.edu/exhibits/rbm/kislakreligion/anonymouss.html
SOLUTION

Advance information
meaning – morphologically segmented:

‘you (SG) are going’  $n$-$at$-$beq$

‘we will go’  $t$-$o$-$beq$

‘you (PL) are going’  $n$-$eš$-$beq$

‘we are ill’  $yaš$-$o$

‘you (SG) are a man’  $gwink$-$at$

‘he is a man’  $gwink$

‘we are women’  $išk$-$o$

‘you (PL) are women’  $išk$-$eš$
Translations – morphologically analysed:

‘I will go’  
\[ t-in-beq \]
FUTURE-1SG-go

‘you (SG) will go’  
\[ t-at-beq \]
FUTURE-2SG-go

‘I am going’  
\[ n-in-beq \]
IN.PROGRESS-1SG-go

‘I am ill’  
\[ yaš-in \]
ill-1SG

Observations:
words for states and properties appear to have the marking for PERSON and NUMBER after stems, words for actions and other processes before stems; also, marking for time (TENSE, ASPECT?) appears to be missing with state or property words (which designate something rather time-stable)
<table>
<thead>
<tr>
<th>English</th>
<th>Syllable</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘he is ill’</td>
<td>yaš(-Ø)</td>
<td>ill(-3SG)</td>
</tr>
<tr>
<td>Observation:</td>
<td></td>
<td>3rd person (singular) – meaning: neither speaker nor addressee – appears to lack overt marking</td>
</tr>
<tr>
<td>‘you (SG) are ill’</td>
<td>yaš-at</td>
<td>ill-2SG</td>
</tr>
<tr>
<td>‘you (PL) are ill’</td>
<td>yaš-eš</td>
<td>ill-2PL</td>
</tr>
<tr>
<td>Observation:</td>
<td></td>
<td>Actually, the minimal meaning difference only concerns number (how many persons are ill): person (addressee) is the same; but there does not seem to be a form contrast only corresponding to the meaning contrast SG – PL, and no number part seems to be segmentable from a person part</td>
</tr>
</tbody>
</table>
‘I am a man’  
\textit{gwink-in}  
\textit{man-1SG}  

‘we are men’  
\textit{gwink-o}  
\textit{man-1PL}  

‘she is a woman’  
\textit{i\texttrans{k}(-\emptyset)}  
\textit{woman(-3SG)}  

Observation:  
Male vs. female appears not to be distinguished  
(for 3SG)  

Source of the Kekché data:  
Ann Arbor: University of Michigan Press.
Now, what exactly have we done when we identified the parts of words in Kekchí, a language where we were unable to draw on our intuitions?

It’s simple:

We **compared** (what? form-meaning pairings) and looked for **identities** and **differences** among the items we compared.

Suppose you only knew this about Kekchí – and note that this is a **form** as well as its **meaning** (What if you didn’t even know the meaning?):

\[ \text{tinbeq} \]  
‘I will go’

Could you work out the morphological word structure?

Well, you could try guess-work. (Based on what? See below, *Expectations.*)
To do systematic morphological analysis, we need several forms-with-meanings, in order to be able to compare and find out something morphologically relevant.

And then we need to compare skillfully/strategically:

We are looking for such (minimal) differences in form which systematically come with (minimal) differences in meaning.
Here now are three forms-with-meanings:

\[ tinbeq \]  ‘I will go’
\[ tatbeq \]  ‘you (SG) will go’
\[ ninbeq \]  ‘I am going (right now)’

Let’s compare each one of them with every other.
tinbeq

tatbeq

_______________________________

same: \( t \ldots beq \)
different: \( . in \ldots / . at \ldots \)

meaning difference:
The subject (the participant actively involved in the event of going) is the speaker (in the case of \(-in-\)) or the addressee (in the case of \(-at-\)).
tinbeq
ninbeq

_______________________________

same: . inbeq
different: t . . . . / n . . . .

meaning difference: The event of going is to take place in the future from the time of speaking (t-) or is in process at the time of speaking (n-).
tatbeq
ninbeq

_______________________________

same: \ldots beq

different: tat \ldots / nin \ldots

meaning differences: (i) The subject is an addressee (tat-) or the speaker (nin-).
(ii) The time of the event is the present (tat-) or the future (nin-), relative to the time of speaking.

That is, there are as many as two meaning differences involved here, hence this is not the wisest comparison to begin the analysis.
(interim) result:

- **beq**  ‘go’
- **n-**  ‘in process’ (simultaneous with time of speaking)
- **t-**  ‘in future’ (after time of speaking)
- **in-**  ‘Subject = the speaker’
- **at-**  ‘Subject = an addressee’
Thus, this is the **word structure**, in terms of meaning categories and the sequence of their expression:

Time — Participant(s): Person — Event

or, more technically:

**TENSE** — Speech-event ROLE (and NUMBER?) of Agent/Subject — Verb Stem

(Possibly **ASPECT** rather than **TENSE**: which are two different, but related categories, not always easy to distinguish. Tense is deictic, aspect is not, and instead differentiates, e.g., completed events vs. events in progress, PERFECTIVE vs. IMPERFECTIVE or PROGRESSIVE.)
And so on, until all words given and their parts are satisfactorily accounted for.

And then we need to test out our hypotheses against further data, going beyond those informing the original analysis. Ultimately we want to be able to account for all of the morphology of our language.

? ‘you (SG) are going’

natbeq

etc.

For those momentarily unable to go off doing fieldwork on Kekché, in Guatemala or wherever you find native speakers (New York City is always a good address for this purpose), there is a homework, “More Kekché”.
Next comes a slightly more complex example from a less unfamiliar language, German – more precisely, Hochdeutsch of a kind where speakers/hearers continue to be able to deal with forms such as *schaute* or *schautest*, thus probably also requiring fieldwork in the far north of the country – to practise this simple method of matching minimal form differences with minimal meaning differences.

Here are the forms to work on – more to be added on demand:

*schaue, schaust, schaut, schauen, schaut, schauen, schaute, schautest, schaute, schauten, schautet, schauten, haue, haust, ..., komme, kommst, kommt, kommen, kommt, kommen, kam, kamst, kam, kamen, kamt, kamen, ...*
Meanings (but then, the meanings of verbs are hard to express through pictures, because pictures typically also depict people, animals, things which are doing what the verbs mean):
or segmentable into \(-st\) ?

Hardly possible to match plausible meaning with \(-s\) and with \(-t\), oder? (2 and SG?)
Starting off the comparing with these two forms, with no other forms and meanings yet available, it would be tempting to identify the minimal form difference as one between the absence and the presence of a final schwa (-e) and to associate this with the minimal meaning difference between present and past time (PRESENT and PRETERITE TENSE).
However, further comparisons will prove this hypothesis untenable.
<table>
<thead>
<tr>
<th></th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ich</strong></td>
<td>schaue</td>
</tr>
<tr>
<td><strong>ich</strong></td>
<td>schaute</td>
</tr>
<tr>
<td>-t</td>
<td>PRETERITE</td>
</tr>
<tr>
<td>-e</td>
<td>Sbj:1SG</td>
</tr>
<tr>
<td><strong>du</strong></td>
<td>schaust</td>
</tr>
<tr>
<td><strong>du</strong></td>
<td>schautest</td>
</tr>
<tr>
<td>-t</td>
<td>PRETERITE</td>
</tr>
<tr>
<td>-(e)st</td>
<td>Sbj:2SG</td>
</tr>
<tr>
<td><strong>er</strong></td>
<td>schaut</td>
</tr>
<tr>
<td><strong>er</strong></td>
<td>schaute</td>
</tr>
<tr>
<td>-t₁</td>
<td>Sbj:3SG.PRESENT</td>
</tr>
<tr>
<td>-e</td>
<td>Sbj:3SG.PRETERITE</td>
</tr>
<tr>
<td>-t₂</td>
<td>PRETERITE</td>
</tr>
</tbody>
</table>
**Word structure** of the German verb, therefore:

Stem (– **TENSE**: PRETERITE) – Subject: **PERSON**.**NUMBER**(.**TENSE**)

\[
schau\- \quad (-t-) \quad -e \quad 1\text{SG} \\
\quad -(e)st \quad 2\text{SG} \\
\quad -t \quad / \quad -e \quad 3\text{SG} \quad \text{PRES} / \quad \text{PRET} \\
\quad -en \quad 1\text{PL} \\
\quad -(e)t \quad 2\text{PL} \\
\quad -en \quad 3\text{PL}
\]

The first parenthesis means: this position can be filled (if **TENSE**: PRETERITE), but does not **have** to be filled (if **TENSE**: PRESENT);

the second parenthesis means: in this position **TENSE** is (co-)expressed, but not always (only in the case of 3SG).
Compare this with the word structure of another type of German verb, where the formal difference between tenses works differently:

<table>
<thead>
<tr>
<th>verb form</th>
<th>morphemes</th>
<th>tense/person</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>komm-</code></td>
<td><code>-e</code></td>
<td>1SG PRES</td>
</tr>
<tr>
<td></td>
<td><code>-st</code></td>
<td>2SG</td>
</tr>
<tr>
<td></td>
<td><code>-t</code></td>
<td>3SG PRES</td>
</tr>
<tr>
<td></td>
<td><code>-en</code></td>
<td>1PL</td>
</tr>
<tr>
<td></td>
<td><code>-t</code></td>
<td>2PL</td>
</tr>
<tr>
<td></td>
<td><code>-en</code></td>
<td>3PL</td>
</tr>
<tr>
<td><code>kam-</code></td>
<td><code>-Ø</code></td>
<td>1SG PRET</td>
</tr>
<tr>
<td></td>
<td><code>-st</code></td>
<td>2SG</td>
</tr>
<tr>
<td></td>
<td><code>-Ø</code></td>
<td>3SG PRET</td>
</tr>
<tr>
<td></td>
<td><code>-en</code></td>
<td>1PL</td>
</tr>
<tr>
<td></td>
<td><code>-t</code></td>
<td>2PL</td>
</tr>
<tr>
<td></td>
<td><code>-en</code></td>
<td>3PL</td>
</tr>
</tbody>
</table>
Lots more exercises in morphological analysis to be found in Nida, *Morphology*. First try the simpler problems, Nos. 4, 5, 6. Ask the tutor for help if you need any.

For the more complex problems in Nida we need something we don’t have yet: the concept of **ALLOMORPHY**.
Reflections

Morphological analysis is learnt by practice; but do occasionally pause and reflect on what you are doing when you are learning by doing.

In this spirit, consider this question.

When we – or actually also language learners, who will need to discover the form-meaning atoms, too, in order to be able to re-combine them in novel word constructions – do morphological analysis and look for minimal differences in form that can be systematically matched with minimal differences in meaning, in Kekchí or any other language, are we really blindly following a mechanical procedure, not influenced or guided by any expectations about the possible results of applying this procedure? Or do we have certain advance assumptions, however well or ill founded, that narrow down what we reckon with as the outcome of our morphological analysis?

Here are some advance assumptions, of a very general kind, that some morphologists (no names named here) may be tempted to subscribe to (tacitly perhaps).
(i) Not only are all conceivable meanings expressible in all languages; but the smallest meaning components are the same or very similar in all languages.

(ii) Meanings are constructed compositionally: the meanings of complex wholes accrue from the meanings of their constituent parts.

(iii) There are different kinds of meanings (e.g., lexical, concrete vs. grammatical, abstract), which are reflected in different kinds of forms (e.g., more vs. less phonological/prosodic weight, conspicuously vs. unconspicuously positioned).

(iv) The same meaning is always expressed by the same form in a given language, and vice versa.

(v) One meaning component will not be split up into separate successive form components; two meaning components will not be conflated in a single form.

Informed by a certain amount of experience with a single language, or with a range of diverse languages, one expects things will not be radically different in yet other languages. And these particular advance assumptions are not implausible, in light of major morphological patterns in many languages, but they are not always unproblematic either – ideals that real-life morphology does not always live up to.
Advance Assumption (i)

The first part of this – The Linguistic Equality Hypothesis – is presumably true: Everything that can be said at all can be said in every language. Naturally, it will be easier to say some things in some languages than in others.

(Some things are unspeakable but writable, at least in certain manners of writing: for example, Tomorrow was Monday. Nobody would say such a sentence, in English or another language. But it is actually attested in several novels which use the technique of “free indirect speech”/Erlebte Rede. You could probably think it, too; but no, not really: I’m sure you'd not think this thought in the past tense. This is an exciting linguistic subject; but not one we can pursue here, despite its morphological implications.)
The second part is also an empirical issue, but one so complex that reliable results aren’t in yet. Therefore, only a few observations here.

In our Kekchí exercise above we used a real language, English, to provide the meanings expressed by Kekchí words. This is easy and could seem innocuous; but there are pitfalls if one expects too much parallelism between the language studied (Kekchí) and the semantic metalanguage (here English).
For example, we hypothesised that Kekchí -o/-o means ‘we’: but what does English we mean?

*We* is known as the 1ST PERSON PLURAL personal pronoun in English, but it can in fact refer to several partly distinct sets of people (whose common denominator is the membership of the speaker):

- the speaker and one addressee;
- the speaker and more than one addressees;
- the speaker and one or several others, not including the addressee(s);
- the speaker and one or several others, including the addressee(s);
- rarely several speakers/writers speaking or writing together.

These are partly distinct meanings, but English does not formally recognise these distinctions. Other languages do, distinguishing a 1st person INCLUSIVE (addressee included) and 1st person EXCLUSIVE (addressee excluded): they would be using a different pronoun for ‘We (you not included) went to the pub last night’ and ‘We (you included) should go now’.

On the evidence given, we don’t know whether Kekchí does; but taking English as our yardstick, we run the risk of not really expecting and therefore overlooking such a meaning distinction.
By the way: Is we really a PLURAL of I, like cats is the PLURAL of cat?

Look up the category of ASSOCIATIVE, which has sometimes been suggested to be more appropriate here: ‘a central referent and other associated with her/him/it’ (e.g., such pater familias and family).
For Kekchí -at/-at we hypothesised the meaning ‘you’, and here we were aware of an English deficiency – namely, not to formally distinguish singular and plural in the 2nd person personal pronoun, in the way German does, du vs. ihr (with the formal pronoun of address Sie again conflating the distinction). We therefore enriched our semantic metalanguage by adding the category labels SG and PL to ordinary English.

But this again goes to show that one ought to reckon with formal contrasts matching up with different meaning contrasts in different languages.

English actually does show a contrast between 2nd person singular and plural, if only in the reflexive or emphatic pronoun, yourself – yourselves. Also, there are contemporary varieties of English which are seeking to recreate new 2nd person plurals such as you-guys, you-all, yous, different from 2nd person singular.
On the other hand, from an English point of view Kekchí is deficient in the 3rd person singular: primed by English *he* vs. *she*, we were probably expecting a gender contrast, masculine/male – feminine/female, but we couldn’t find one. Should we continue searching?

Come to think of it, there were no overt forms for 3rd person in Kekchí at all, other than bare stems: Should we still set up inaudible segments (Ø, zero or null) expressing this meaning, which after all is the only one overtly expressed in English verbs (*I go, you go, he/she/it go-es, we go, you go, they go*)?
In this domain of pronouns, it looks like the smallest meaning components are not necessarily the same in all languages, then – if we only recognise a meaning contrast if it is supported by a form contrast in the language concerned.

On the other hand, if you survey those categories that are expressed through morphology (especially inflectional morphology) across the languages of the world, it is striking that the same categories tend to occur again and again: CASE, NUMBER, PERSON, DEFINITENESS, GENDER, POSSESSION, TENSE, ASPECT, MOOD, EVIDENTIALITY, VERBAL VOICE, COMPARATIVE, SUPERLATIVE, and perhaps a few more. There won’t be many surprises, really. Well, there are SWITCH REFERENCE, LOGOPHOR, OBVIATIVE/PROXIMATIVE, DIRECT/INVERSE, SUPINE, CONATIVE, MIRATIVE ...

In one theory, that of “Natural Semantic Metalanguage” as advocated by Anna Wierzbicka & Cliff Goddard in many publications (but it is not uncontroversial – what is), the claim has been made that there is indeed a finite list of universal semantic primitives – namely these (http://en.wikipedia.org/wiki/Natural_semantic_metalanguage):
## Proposed and Experimentally Supported Semantic Primes

<table>
<thead>
<tr>
<th>Category</th>
<th>Primes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Substantives</strong></td>
<td>I</td>
</tr>
<tr>
<td>Relational Substantives</td>
<td>SOMETHING/THING</td>
</tr>
<tr>
<td>Determiners</td>
<td>THIS</td>
</tr>
<tr>
<td>Quantifiers</td>
<td>ONE</td>
</tr>
<tr>
<td>Evaluators</td>
<td>GOOD</td>
</tr>
<tr>
<td>Descriptors</td>
<td>BIG</td>
</tr>
<tr>
<td>Mental/Experiential Predicates</td>
<td>THINK</td>
</tr>
<tr>
<td>Speech</td>
<td>SAY</td>
</tr>
<tr>
<td>Actions And Events</td>
<td>DO</td>
</tr>
<tr>
<td>Existence And Possession</td>
<td>THERE IS/EXIST</td>
</tr>
<tr>
<td>Life And Death</td>
<td>LIVE</td>
</tr>
<tr>
<td>Time</td>
<td>NOW</td>
</tr>
<tr>
<td>Space</td>
<td>WHEN/TIME</td>
</tr>
<tr>
<td>Logical Concepts</td>
<td>WHERE/PLACE,</td>
</tr>
<tr>
<td>Intensifier, Augmentor</td>
<td>SIDE</td>
</tr>
<tr>
<td>Similarity</td>
<td>NOT</td>
</tr>
<tr>
<td></td>
<td>VERY</td>
</tr>
</tbody>
</table>
Advance Assumptions (iv) and (v)

The ideal sign (occasionally/often reality)

\[
\text{FORM} \quad \text{MEANING}
\]

Deviations from the ideal of bi-uniqueness

\[ F \]
\[ M_1 \quad / \quad M_2 \]

Ambiguity

Homonymy:
Grm Bank – Bank
-t PRET – -t Sbj:3SG PRES in German verb inflection
Engl no – know,
-s Sbj:3SG in verb inflection – -s PL in noun inflection

Polysemy:
Grm Fuß
Grm wer interrogative, (free) relative, indefinite pronoun
Synonymy

Engl *lorry* – *truck*; *buy* – *purchase*
Grm PL -s (as in *Oma-s*) – -en (as in *Frau-en*) in noun inflection
Engl [hɪm] – [hɪmn] (as in *hymn* – *hymn-ic*)

Cumulation

Engl *punch, slap* … (see below)
Grm -e Sbj:3SG.PRETERITE in verb inflection
Lat -mini Sbj:2PL.IND.PRES.PASSIVE in verb inflection

Extension

Lat *mo-mord-i*
  PERF-bite-1SG.IND.PERF.ACT
Grm *ge-schau-t*
  PARTICIPLEII-look-PARTICIPLEII
**Advance Assumption (ii)**

Is *Hand-schuh* ‘glove’ compositional? (= meaning of the whole a function of the meaning of its parts; and *Hand* and *Schuh* would really seem to be parts – something you’d want to segment when doing morphological analysis!)

*Bahn-hof?*

*Hänge-matte?*

*Arm-brust?*

*Männ-chen?* (in the zoologists’ sense, who can thus non-affectively refer to quite large an animal, like an elephant)

*be-greifen?*

*be-kommen?*

*ver-stehen?*
Where is non-compositionality less unexpected, in word-constructions (morphology) or in phrase- or clause-constructions (syntax)?

[Answer: the former. But don’t forget idioms, such as kick the bucket, a verb phrase consisting of verb and object, the object a noun phrase consisting of definite article and noun – a syntactic construction that is non-compositional.]

Why?

[Answer: Because words are likelier to be memorised than phrases or whole clauses and sentences. What is individually stored in memory, rather than constructed on demand, is likelier to individually undergo further changes than what is again and again constructed in accordance with general rules.]
Advance Assumption (iii)

To come back to our introductory examples from English:

Would /ð/ and /ə/ as in the, both just a single segment, be too short, phonologically too insubstantial, to carry meaning all on their own?

Well, let’s analyse, rather than opine.

Where does English have /ð/ as an onset?

(Not confuse with /θ/, voiceless, as in theory, theme, thesis, thin, thigh, thing, thistle, through, three, thug, thwart, etc.!)  
the₁, this, these, that, those,  
they, them, their, theirs, (thou, thee, thy, thine)  
there, thither, therefore, thus, then, thence,  
than, though, the₂ (as in the sooner the better)

These are all grammatical words, often local demonstratives, or at any rate definite.
Does /ð-/ mean DEFINITE?

What would the remaining part, /ə/, mean, then?

- It is a form with a meaning which contrasts with the meaning of -is in *this*, -at in *that* etc.
- It is a form which occurs without *th-* in *a mat* /ə.'mæt/ and here has the same meaning as in combination with *th-*
  (Problem: *the idea* /ði.ˈaɪə/ — *an idea* /ə.ˈnɪə/)
On the expectation that formal insubstantiality militates against morph status:

- Latin *dedisse:mus* ‘we would have given’
  - the 3rd sound segment, /d/, all on its own carries the meaning ‘give’
  (Latin is an Indo-European language, subfamily Italic; not a “dead” language, strictly speaking, but continuing life in today’s Romance languages.)

- Georgian
  
  *m-e-ši-n-i-a, vai-tu rus-eb-i*
  
  I-IOV-afraid-PRES-it, that Russ-PL-NOM
  
  *še-mo-gv-e-s-ev-i-an(=me+tk+i)*
  
  PREV-PREV-us-IOV-attack-TS-IND.FUT-they(=SP.PART)
  
  ‘I fear that the Russians will attack us’

  - the verb form consists of 8 syllables; its stem is the onset of the fourth syllable
  (Georgian is a Kartvelian, or South Caucasian, language.)
BPG Glaхо: ქართული სიტყვები იხსოვს გამახვიდული და მართლწმინდი სიდუმლით და ელემენტებს. მათ მხრიდან ქართული ნამუშევარი და სისტემა და გრაფიკური სიმბოლო ფუნქცია უთქვა უძლოვალი შიგნიში სახელურება.

BPG Chveulebrivi: ქართული სიტყვები იხსოვს გამახვიდული და მართლწმინდი სიდუმლით და ელემენტებს. მათ მხრიდან ქართული ნამუშევარი და სისტემა და გრაფიკური სიმბოლო ფუნქცია უთქვა უძლოვალი შიგნიში სახელურება.

BPG Courier: ქართული სიტყვები იხსოვს გამახვიდული და მართლწმინდი სიდუმლით და ელემენტებს. მათ მხრიდან ქართული ნამუშევარი და სისტემა და გრაფიკური სიმბოლო ფუნქცია უთქვა უძლოვალი შიგნიში სახელურება.
Despite such cases where concrete lexical meanings require minimal formal effort for their expression (a single sound segment!) and abstract grammatical meanings contribute the lion’s share of the form of words, matters tend to be as expected from advance assumption (iii):

- **morphs** (especially lexical ones) tend to be at least a **syllable** long;
- **lexical** morphs tend to be more substantial than **grammatical** morphs.
SYLLABLE ≈ MORPHEME / WORD

cf. English
have, do, be, see, hear, feel, smell, taste, eat, drink, chew, run, walk, swim, fly, kick, kill, miss, hit, strike, punch, cut, bend, hold, blow, guess, love, hate, like, own, get, give, send, cast, throw, pay, sell, buy, speak, talk, cry, shout, yell, read, write, mean ...

but cf. also:
linguistic(s), enamel, tambourine, phenomenon, imbroglio, hippopotamus, rhinozeros, symposium, mulligatawny, pumpernickel, sauerkraut, humdinger, Potawatomi ...

(Are 5 syllables the limit for English? Is there also a lower limit for the sound substance of words? Like this: A word must minimally consist of a heavy syllable. Notice that none of the examples above consists of only a light syllable. Can you find examples of such super-light words from languages you are familiar with?)
For which of the Advance Assumptions are the following English data a problem?

\begin{itemize}
  \item \textit{hit} ‘bring hand or something held in hand forcefully in contact with surface of someone or something’
  \item \textit{strike} ‘hit sharply or forcefully’
  \item \textit{kick} ‘hit with foot’
  \item \textit{punch} ‘hit with closed hand (fist)’
  \item \textit{slap} ‘hit with flat part of open hand’
  \item \textit{smack} ‘hit with flat part of open hand, making a loud noise’
  \item \textit{spank} ‘hit with open hand, esp. on buttocks’
  \item \textit{butt} ‘hit/strike/push with head or horns’
  \item \textit{nudge} ‘hit/push with elbow’
\end{itemize}
the meanings are complex, and all are complex in the same way
(‘to hit, = bring forcefully in contact with surface of someone or something’ + ‘with a body part’), but there is no parallel formal complexity; hence no morphological segmentation is possible

(Such sets of words form what is sometimes called a “lexical field”).
Further English stuff to practise morphological analysis: Does *mattress* contain *mat*?

cf.  
- *add*  
- *waiter*  
- *tiger*  
- *master/mister*  
- *murderer*  
- *seam*  
  ‘a line of stitches joining two pieces of cloth, leather, etc.’
- *fort*  
- *butt*  
  ‘to strike or push the head or horns against s.o./s.th.; a push’;  
  ‘a person or thing that people make fun of’;  
  ‘a large, thick, or bottom end’

*address*  
*waitress*  
*tigress*  
*mistress*  
*murder[er]ess*  
*seamstress*  
*fortress*  
*buttress*  

looked at historically:

E *mattress* < Old French *materas* < Arabic *matrah* ‘where something is thrown down; where one lies down’
http://www.cartoonstock.com/directory/c/carpet_shop.asp
Morphology or No Morphology? – that is the question

- What is the problem of cranberries, mulberries, huckleberries, cobwebs, scapegoats, twilight, hinterlands, ruthlessness, inertness, ...?

Is cranberry segmentable into cran-berry?
No problem concerning berry, which occurs independently and in other combinations. But the part that is left over, cran: What is its meaning?
Always the same meaning, wherever it occurs?
But where else does cran occur, on its own or in combinations?
Nowhere (in contemporary English): that’s why such forms are called UNIQUE or CRANBERRY morphemes.
Cranberries (Vaccinium macrocarpon)
Arguably, they do have meaning, though: *cran* contributes to *cranberry* what distinguishes the meaning of *cranberry* from the meaning of *strawberry*, *blackberry*, etc., as well as from the meaning of *berry* on its own.

Same reasoning for the other examples: *mul*, *huckle*, *cob*, *scape*, *twi*, *hinter*, *ruth*, *ert*.

If there are speakers of English who can associate the *cran* in *cranberry* with *crane* (the bird, German *Kranich*, or also the machine for moving heavy objects, German *Kran*) or who can associate the *scape* in *scapegoat* with *escape*, etc., then these are not unique morphemes for such speakers.

But the mental lexicons-and-grammars of such speakers will then also need some further extras: they need to account for the (unique) formal alternations involved, *[kræn]* – *[kreɪn]* and *[skeɪp]* – *[ɪˌskeɪp]*; and they need to provide plausible comprehensive meanings covering all occurrences of these supposed non-unique morphemes.

(There is an extra text on the morphological peculiarity of designations of berries.)
cranberry crane  flower of *Vaccinium macrocarpon*  crane (*Grus grus*)
• What is the problem of *flow*, *flare*, *flitter*, *glow*, *glare*, *glitter* ...?

Intuitively, it was conjectured above, one might be tempted to segment such words into the parts *fl/-fl/**, *gl/-gl/**, *-ow/-ow/**, *-are/-e(r)**, *-itter/-t(r)**.

Going about it methodically, one can indeed identify minimal form differences that can be matched with minimal meaning differences – meanings which are to do with sensory experience and which may perhaps be a bit general and/or hazy:

- *fl/**: ‘movement’
- *gl/**: ‘light, shine’
- *-ow**: ‘steady’
- *-itter**: ‘unsteady’
- *-are**: ‘intense’

These form-meaning parts are not like *cran-*: they do combine with more than one other form-meaning part, contributing the same meaning. (Hence our tentative intuitions.) But they do not re-combine prolifically. It is easy to construct complex meanings with components such as ‘movement’, ‘light’, etc., although without being able to use these forms as constituents of corresponding complex forms.
We will find **limited productivity** also elsewhere in otherwise impeccable morphology, especially the branch known as “word formation”. What is special about the present cases is the **unaccounted left-overs**. This is to say, it is not difficult to cite further words which, with a little imagination, can be analysed as containing a form-meaning matching in question – for example:

‘movement’:   fl-y, fl-it, fl-ick, fl-oat, fl-ush, fl-ap, fl-ip, fl-op, fl-ee, fl-ing, fl-unk, ...
‘light, shine’:   gl-isten, gl-eam, gl-int, gl-immer, gl-impse, gl-ass, gl-ance, gl-oss, ...

But what about the remainders of these words? Can the word-ends -y, -it, -ick, -isten, -eam, -int, etc. also be matched up with an appropriate meaning? With meanings which these form parts would also express in other combinations? Hardly. (Hence the tentativeness of our intuitions.)

Now then, what to do with *fl-, gl-, -ow, -are, -itter* and their ilk?
- Ignore them in a scholarly analysis of the structure of complex words in English, as figments of the imagination of speakers who aren’t morphologists?
- Or find some special niche for them, as units distinct from morph(eme)s, though akin to them? Let’s call them “submorphemic differentials” or “phon(a)esthemes”, though a rose by any other name would smell as sweet.
Read these advocates of the second position:

Marchand, Hans. 1959. Phonetic symbolism in English word-formation. *Indogermanische Forschungen* 64. 146-168. (Also Marchand’s handbook of English word formation.)

Further relevant pointers:  ICONICITY, IDEOPHONE
• **What is the problem of English receive?**

Is it segmentable into *re-ceive*? Does *ceive* have meaning? Always the same meaning wherever it occurs?

Unlike cranberry morphemes *ceive* does occur in other combinations, at least three:

\[
de-ceive, \; per-ceive, \; con-ceive
\]

It isn’t easy to tell what this /siv/ could mean in these four English verbs. A meaning ‘grasp’ or ‘take’ – knowing Latin, this is what one might want to go for – is not really appropriate for *de-ceive*. And which English speakers all over the world who have these verbs in their mental lexicons know Latin?

But it is also easy to see that these four /siv/’s have more in common than this particular sound form.
First, the words they form part of are all four of them **verbs**. In particular, they are all **transitive** verbs (verbs requiring a direct object).

Second, /siv/ always **changes its shape** in exactly the same way when these verbs are combined with further form-meaning components – such as -tion, which transforms verbs into nouns and /siv/ into /sɛp/: **re-cep-tion, de-cep-tion, per-cep-tion, con-cep-tion**.

This is to say: **ceive** in **re-ceive** etc. may not always have one and the same **meaning** (whatever that meaning would be), as ordinary morph(eme)s do, but it always has the same **grammar**.

To be able to extend the notion of morph(eme) to such creatures, one would have to define **morph(eme)** as the **smallest form-unit** which has a (distinctive) **meaning** and/or a (distinctive) **grammar**.
Further homework

Does one really always know intuitively whether a word in a familiar language is morphologically segmentable?

We’ve already discussed two English problem cases: *th-e* and *mat(t)-ress*.

How about these: *E lord, lady, female* – complex or simple?

Here we’re talking about today’s English, the stage of the language we’re most likely to be familiar with. What about earlier stages, such as Middle or even Old English? Check out a suitable dictionary (that is, an etymological dictionary; but the grand *Oxford English Dictionary* (OED) also provides historical information) whether something relevant has changed here.
2.2. **CLASSIFICATION** of minimal form-meaning segments (= morphs)

Meaning/grammar-ful word-parts are not all of the same kind: there are several *classes* of them, depending on which criteria they are classified by.

Not only morphologists can classify morphs; this is something everybody can do who has some intuitive understanding of their own language: morpheme classes are part of mental grammar, not just of grammar books by linguists.

Speakers of **Golagat**, for example, are expert morph segmenters and classifiers. Golagat is Tagalog spoken backwards, a word play children in the Philippines like to play – where Tagalog is one of the official languages, among some 70 non-official languages also spoken there.

Tagalog → Golagat

Bading  Ngidab  ‘Bading’ (a personal name; ng is apparently considered a single sound, because it isn't reversed)

lima  amil  ‘five’
laki  ikal  ‘big’
sulat  talus  ‘to write’
kain  niak  ‘to eat’ (or perhaps rather 'to have eaten'; ai is apparently considered to be two sounds)

And here a whole sentence, with each word reversed, but the sequence of words the same:

Upo ka na muna dito  →  Opu ak an anum otid

‘sit you now a.while here!’
And here are some further reversals: Aren’t they puzzling?

\[
\begin{align*}
\text{tiglima} & \quad \rightarrow \quad \text{tigamil} \quad & \text{‘five each’} \\
\text{sulatin} & \quad \rightarrow \quad \text{talusin} \quad & \text{‘to be written’} \\
\text{kumain} & \quad \rightarrow \quad \text{numiak} \quad & \text{‘(he/she/it) ate’} \\
\text{kakain} & \quad \rightarrow \quad \text{niniak} \quad & \text{‘(he/she/it) will eat’} \\
\text{kumakain} & \rightarrow \quad \text{numiniak} \quad & \text{‘(he/she/it) is eating’}
\end{align*}
\]

Not really. Do it yourself:

\[
\begin{align*}
\text{malaki} & \quad \rightarrow \quad \underline{\hspace{2cm}} \quad & \text{‘to be big’} \\
\text{pumasok} & \quad \rightarrow \quad \underline{\hspace{2cm}} \quad & \text{‘(he/she/it) entered’} \\
\text{sasama} & \quad \rightarrow \quad \underline{\hspace{2cm}} \quad & \text{‘(he/she/it) will come’} \\
\text{pumapasok} & \quad \rightarrow \quad \underline{\hspace{2cm}} \quad & \text{‘(he/she/it) is entering’}
\end{align*}
\]
What exactly do Golagat-speaking kids know about morph-classification?

They know that morphs come in two classes.
They are aware of the distinction between STEMS and AFFIXES.

It is as if in speaking Deutsch Rückwärts (Can you?) you’d reverse as follows:
\[ be\text{-}strumpf\text{-}st \rightarrow be\text{-}pfmursch\text{-}st, *tspfmutrscheb \]
But then, German is not easy to speak backwards in the first place. (Why?)

First analyse the examples yourselves. Then look at the extra file on the subtleties of Tagalog morphology that Golagat speakers have no difficulties to master, hard nuts though they are for morphologists.
Criteria for morph(eme) classification:

- **by status**
  - free, lexical or bound, grammatical
  - = word, stem, root = affix or other exponent
  - LEXEME GRAMMEME

  e.g., Kekchí
  - *-beq ‘go’*
  - *n- ‘in progress’*
  - *t- ‘future’*
  - *-in- ‘Subject = speaker’*
  - *-at- ‘Subject = one addressee’*
e.g., German

`schau-` ‘intentionally perceive with one’s eyes’

- \textit{t-} PRETERITE
- \textit{e} Sbj = 1SG
- \textit{(e)st} Sbj = 2SG
- \textit{t} / \textit{e} Sbj = 3SG PRES / PRET
- \textit{en} Sbj = 1PL
- \textit{(e)t} Sbj = 2PL
- \textit{en} Sbj = 3PL

“Free” is not always to be taken entirely literally. In the case of \textit{schauen} it is true that the stem can occur on its own in syntactic constructions: the bare stem forms the 2SG imperative, \textit{schau!}.

On the other hand, \textit{-beq} ‘go’ in Kekchí or \textit{d-} ‘give’ in Latin cannot be used syntactically without one or the other affix.

In German, a noun stem such as \textit{Kirch-} also requires affixation, through a stem formative \textit{-e}, when used in constructions or when the word is cited.
• by meaning

• forms designating persons, things, actions, properties, etc. (general criterion for such distinctions: time stability?)
  – and this is a distinction that matters for grammar: word classes!
  noun = Dingwort, verb = Tunwort, adjective = Eigenschaftswort, adverb = Umstandswort

• forms designating something concrete vs. abstract, particular vs. general
  – not the most tangible distinctions perhaps, but it does make sense to say, does it not, that the meanings of Kekchí -beq or German schau- (= lexemes, stems) are comparatively speaking more concrete and particular, less abstract and general, than the meanings ‘being in progress’ (PRESENT, PROGRESSIVE), ‘to occur in future’ (FUTURE), ‘Subject = speaker’, ‘Subject = an addressee’ (= grammemes, affixes)
• by position
  • closer to stem/more distant from word edge
    or more distant from stem/closer to word edge
    
    cf. un-de-cipher-abil-ity, re-place-ment-s, anti-dis-establish-ment-ar-ian-ism

  • before, after, inside, around the stem
    (prefix, suffix, infix, circumfix)

Note: -ment and -s are both suffixes, because both are placed after the stem
– one closer to the stem, the other closing off the word at its right edge;
-ment is no infix, even though it is inside another suffix – its position is after
the stem, not inside it.
• yet further classifications? Yes: to come later