Tagmemics

An Introduction to Linguistics for Perl Developers

or

"Wouldn't know a tagmeme if it bit me on the parse."

Allison Randal
University of Portland
YAPC::EU 2002
In the Beginning...

Wants pawn term dare worsted ladle gull hoe lift wetter murder inner ladle cordage honor itch offer lodge, dock, florist. Disk ladle gull orphan worry putty ladle rat cluck wetter ladle rat hut, an fur disk raisin pimple colder Ladle Rat Rotten Hut...

- Howard L. Chace, *Anguish Languish*
Why Linguistics?

• Linguistics and Software Development?

• Linguistics is the study of language.

• Programming languages are human languages.

• It's the same brain.
Why Tagmemics?

• Out of hundreds of theories, why this one?

• Understand Larry.

• Know your roots.

• Tagmemics is a practical theory.
Who's Pike?

- Dr. Kenneth L. Pike, 1912 - 2000
- Professor of Linguistics at the University of Michigan.
- President of the Linguistic Society of America and SIL International.

© 2002 Allison Randal
Who's Pike?

- Nominated for the Nobel Peace Prize on 15 consecutive years.
- Also known for his entertaining stories, poetry and songs and his blue cape.
- He would fit right in the Perl community.

© 2002 Allison Randal
Where and When

- Pike developed tagmemics while doing field research and teaching linguistics.
- He wanted a theory that was easy to learn and easy to use...
- ...but complex enough to explain real language.
- The Swiss-Army knife of linguistic theories.
Language in Context

- Language cannot be separated from the human context in which it is used.
- Many linguistic theories are satisfied with structure alone.
- Those theories have no explanation for:
  
  *Wants pawn term dare worsted ladle gull...*
Any element can be analyzed in one of 3 ways:

- Particle: each element is a discrete unit.
- Wave: elements are defined by a nucleus.
- Field: relationships between elements are what matters.
The 3 analyses applied to "versions of Perl":

- Particle: each version stands on its own.
- Wave: a release is a nucleus, but development is continuous.
- Field: versions contrast by their sets of features.
What's a Tagmeme?

• A tagmeme is a unit-in-context.

• Tagmemes are fractal.

• A tagmeme has 4 parts: slot, role, class and cohesion.

• Don't get too tied up in the details...
What's a Tagmeme?

Slot | Class
Role | Cohesion
<table>
<thead>
<tr>
<th>Role</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>What's a Tagmeme?</td>
<td></td>
</tr>
<tr>
<td>Where does the unit go?</td>
<td>What kind of unit is it?</td>
</tr>
<tr>
<td>Why is this unit here?</td>
<td>How does this unit relate to other units and to the wider context?</td>
</tr>
<tr>
<td>What's it's function?</td>
<td></td>
</tr>
</tbody>
</table>
The monkey bit me.

What's a Tagmeme?

- Margin
- Subject
- Actor
- Noun Phrase

Obligatory Number agreement with the verb

© 2002 Allison Randal
What's a Tagmeme?

<table>
<thead>
<tr>
<th>Margin</th>
<th>Pronoun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>Obligatory Number agreement with the verb Case marking</td>
</tr>
<tr>
<td>Actor</td>
<td>He bit me.</td>
</tr>
</tbody>
</table>

© 2002 Allison Randal
What's a Tagmeme?

Nucleus
Predicate

Verb

Statement

Transitive
Number agreement with the subject
Case marking

The monkey *bit* me.
What's a Tagmeme?

<table>
<thead>
<tr>
<th>Margin</th>
<th>Pronoun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object</td>
<td>Obligatory Case marking</td>
</tr>
</tbody>
</table>

The monkey bit me.

© 2002 Allison Randal
What's a Tagmeme?

Nucleus
Predicate

Verb (function)

Imperative

Obligatory

print $animal;
What's a Tagmeme?

<table>
<thead>
<tr>
<th>Margin</th>
<th>Noun (variable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object</td>
<td>Optional</td>
</tr>
</tbody>
</table>

Undergoer

print $animal;
What's a Tagmeme?

Margin
Object

Noun (string)

Undergoer

Optional

print "antelope";
What's a Tagmeme?

Role
Class

Slot

Object

Undergoer

Statement

Optional

print join("\t", @animals);
Etic and Emic

• Two different perspectives on the same data.

• Etic is logical, alien, external, invented scientifically measurable.

• Emic is relative, native, internal, discovered, what matters to participants.
Etic and Emic

- Etic is E.T.-ic.
- Emic is ME-ic.
• An etic view of a television:
  – vaguely cubical, but irregular shape
  – dark color
  – dimensions range from inches to yards
  – one surface lighted with changing colors
  – emits sound
An emic view of a television:
- the colors represent human activity
- a connection to the world
- or an escape
- an education tool
An etic perspective on language design:
– X number of control structures...
– X number of forms to express them.
Etic and Emic

• An emic perspective on language design:
  – Who is going to use it and how?
  – What other concepts are they going to be familiar with?
  – What will they learn easily?
  – What will be meaningful?
  – How will it relate to other aspects of the language?
Re-analysis

```c
switch (c) {
    case '0':
        zero++;
        break;

    case 'a':
        alpha++;
        break;

    default:
        nomatch++;
        break;
}
```
```
switch (c) {
    case '0':
        zero++;
        break;

    case 'a':
        alpha++;
        break;

    default:
        nomatch++;
        break;
}
```
switch (c) {
    case '0':
        zero++;  
        break;
    case 'a':
        alpha++; 
        break;
    default:
        nomatch++; 
        break;
}
given (c) {
    case '0':
        zero++;  
        break;
    case 'a':
        alpha++;  
        break;
    default:
        nomatch++;  
        break;
}
given (c) {
  when '0':
    zero++;
    break;

  when 'a':
    alpha++;
    break;

  default:
    nomatch++;
    break;
}

© 2002 Allison Randal
given (c) {
  when '0':
      zero++; break;

  when 'a':
      alpha++; break;

  default:
      nomatch++; break;

}
given (c) {
    when '0' {
        zero++;  
        break;
    }
    when 'a' {
        alpha++; 
        break;
    }
    default {
        nomatch++; 
        break;
    }
}
given (c) {
    when '0' {
        zero++;  
        break;
    } 
    when 'a' {
        alpha++;    
        break;
    } 
    default {
        nomatch++;    
        break;
    } 
}
given (c) {
    when /glick/ {
        zero++;  
        break;
    }
    when 'a' {
        alpha++;  
        break;
    }
    default {
        nomatch++;  
        break;
    }
}
given (c) {
    when /glick/ {
        zero++;
        break;
    }
    when Acme::Snark {
        alpha++;  
        break;
    }
    default {
        nomatch++;  
        break;
    }
}
Re-analysis

given (c) {
    when /glick/ {
        zero++; break;
    }
    when Acme::Snark {
        alpha++; break;
    }
    default {
        nomatch++; break;
    }
}
Re-analysis

given (c) {
    when /glick/ {
        zero++;
    }
    when Acme::Snark {
        alpha++;
    }
    default {
        nomatch++;
    }
}
given (c) {
    when /glick/ {
        zero++;
    }
    when Acme::Snark {
        alpha++;
    }
    default {
        nomatch++;
    }
}
given ($c) {
    when /glick/ {
        $zero++;
    }
    when Acme::Snark {
        $alpha++;
    }
    default {
        $nomatch++;
    }
}
Designing a language without considering that people are going to use it is about like designing a car without considering that people are going to drive it.
Further Reading


*(Both titles are out of print, but they are available at several used book sellers online.)*
Provides a comprehensive introduction to the field of contact linguistics. Examines a wide range of language contact phenomena from both general linguistic and sociolinguistic perspectives. Offers an account of current approaches to all of the major types of contact-induced change. Discusses the general processes and principles that are at work in cases of contact. About the Author. Donald Winford is Professor of Linguistics at Ohio State University. He is author of Predication in Caribbean English Creoles (1993), and co-editor of several volumes, including Focus and Grammatical Relations in C An Introduction to Language has been added to your Cart. Add to Cart. Turn on 1-Click ordering. This book presents a concise introduction to natural language processing both theoretical and practical in nature with emphasis on English, French and German language structure. The text is suitable for researchers and students of natural language processing and computational linguistics. In fact, there is an elaborate internet site dedicated to this book. Pierre Nugues has taught natural-language processing and computational linguistics at the following institutions: ISMRA, Caen, France; University of Nottingham, UK; Staffordshire University, UK; FH Konstanz, Germany; Lund University, Sweden; and Ghent University, Belgium. Read more. Product details. It offers a good basic introduction to linguistics if you don’t have much of a background in that. I think it’s pretty standard reading in introductory computational linguistics classes. However, it might still be too theoretical if what you want to do is basically learn how and when to use specific tools or very specific approaches. Oxford University Press. Hudson, Richard. 2010. An Introduction to Word Grammar. Cambridge University Press. And if one is particularly willing to invest time and energy in establishing a solid base in DG and its use for computational goals, then this conference would allow one to connect to the international DG community: http://depling.org/depling2015/.