

BLOCK THREE: Content Analysis

- Overlap between:
 - Content Analysis and
 - Qualitative analysis,
 - text-management,
 - "natural language processing" and
 - some artificial-intelligence software.

BLOCK THREE: Content Analysis

- CA is **top-down** in the sense that:
 - Set of pre-existing categories (“Dictionary”) and specification (instances) are applied to the text
 - Usually by attaching (semantic, grammatical) **tags** to the text material
 - This interpreted material is then typically **counted** within the [user-specified] context, to give rise to a **profile** for each text unit.

BLOCK THREE: Content Analysis

Original versions

- History back to 18th century ... hymns
- Classic accounts :
 - Stone et al, 1966; Krippendorff, 1980; Berelson, 1971; Shapiro and Markoff, 1997 (see Sites/Refs)
- Extensively used in communications research,
 - With a view to:
 - Summarizing,
 - Lexical data-reduction,
 - Distillation of meaning / themes

BLOCK THREE: Content Analysis

Original versions

- Originally, CA Dictionary was:
 - Specified by theorist/researcher to implement theoretical schema / themes
 - Tagging was performed by human coders
 - Often graduate students ...or academic wives,
 - working to a Coding Rule-book
 - To identify **occurrences** of themes
 - With emphasis on **reliability (inter-coder)**
 - Not only to textual material but other cultural artefacts
 - E.G. McClelland The achieving society; Pottery styles ...

BLOCK THREE: Content Analysis

Computer-assisted CA

- With advent of computer, prospect of routine coding as automated process.
- The General Inquirer (Phil Stone, Harvard, 1965)
 - CA: “any technique for making inferences by objectively & systematically identifying specified characteristics of messages”
 - <http://www.wjh.harvard.edu/~inquirer/>
 - ... and Harvard-Edinburgh-Sydney 1970!
 - Maps (large) textual data-sets onto EXISTING [hence ‘top-down’] theoretical category-system (currently Harvard IV dictionary) to produce a **profile** for each textual unit

BLOCK THREE: Content Analysis

INQUIRER III -- Components

- **Text preparation/Reader**
 - N.b. No “...” elision; (*this is comment*); * as EoR
- **Application of Dictionary Categories to text**
 - Lemmatising to reduce to word-stem
 - Disambiguation to differentiate word-senses for homographs
 - Tagging: each word(-sense) in sequence looked up in H-IV and relevant category tags appended (I.e. multiple tag allocation)
- **Analysis:**
 - Profile count for text units (-> MDS, Corresp. An)
 - TRIC (“Tag Retrieval in Context”)

GENERAL INQUIRER ...

- maps each text/segment with counts on 182 dictionary-supplied categories
 - Harvard IV Psycho-social dictionary categories
 - 5 categories of Semin and Fiedler (social cognition)
 - Lasswell (politics) dictionary
 - Each **CATEGORY** = list of words and word senses
 - smallest: “self-references” 12 pronouns
 - largest “negative” 2291 entries
 - Users can add category/ies
- Most of “work” of Inquirer is:
 - Lemmatising (stem reduction)
 - Disambiguation [restricts it to English]
 - Designed for large texts, with speed: can assign the 182 categories to text at rate of several million words/hour

BLOCK THREE: Content Analysis

INQUIRER III Dictionary

- **H-IV Dictionary** is a category system*, based
 - primarily on the **Action Frame of Reference** (sfa)
 - Talcott Parsons, Edward Shils, RF Bales ...
 - Lasswell Value Dictionary (Politics)
 - Osgood's Semantic Differential Factors
 - Includes c182 categories; and over 13K word(-sense)s – "entries" or "instances" and 6.3K disambiguation rules
 - E.g.: Institutional context; Roles; Processes of Perception; Stylistic (*understate*)
 - Is extendable (with great care: see disambiguation)
 - *See *.../homecat* in Website ref (3a) for full specification of category tags

Lemmatize

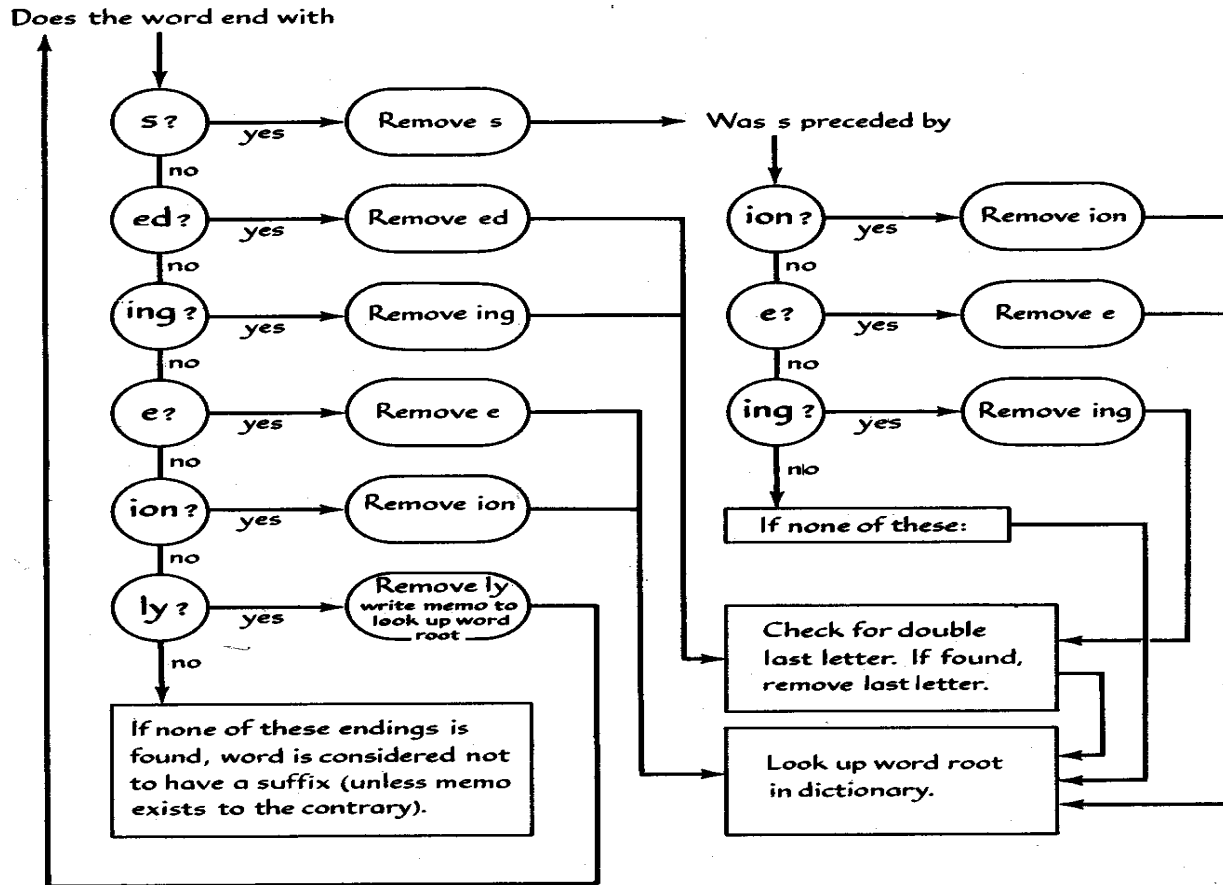


FIGURE 3.9. A suffix-editing procedure for English.



Inquirer Categories and Thematic Analysis

- Inquirer categories (or subset) can be used to produce count Profiles of textual units
 - questions; sections/segments; speakers, texts
 - i.e. use multiple categories [of GI] and analyze the tag patterns assigned to each sentence/text.
- Inquirer Categories (H-IV dictionary entries) and defining “instances” (word-senses) can be extracted (“cannibalised”) and used in HAMLET &c to increase comparability ... even (God forbid!) modified!

Inquirer Categories and Thematic Analysis

- **THEMES** are often a species of **Predication** ... but they can also be defined in terms of concatenations of **categories**:
- (Cf Definition of “STATUS” as a theme in POOC; (C&H p75) *in terms of* Harvard IV Socio-Psychological Dictionary tags):

“Status” as Theme in POOC

THEME 1: STATUS (1)

Key terms: *Profession(s)* (NOUN.MASS.COLL.ECON.ECON.⁵)

chosen-, business-, caring-,

Professional (MODIF.EVAL.ECON.ECON.HUMAN)

- group, - man, - men, - person, - people, - classes,
- job
- interests, - qualifications, - standards, - qualities

Status (=STATUS 1: NOUN.ECON.POLIT.THINK.EVALU.MEANS.POWER)

job -

- in community

also (*Social*) position, looked up to, (*Social*) scale