WordMinerと他の応用ソフトの併用

—内容分析の観点から—

大阪大学大学院（日本学術振興会）
樋口 耕一

概要

- テキストマイニングと内容分析
  - 内容分析におけるコーディング
  - 量的方法と質的方法
  - テキストマイニングにコーディングを取り入れる / 接合する
- WordMinerとコーディングソフト（KH Coder / 茶筌）の併用
  - WordMiner・KH Coderの操作手順
  - 分析事例
    ※必要となる操作手順を織り交ぜながら、分析事例を当日示す

内容分析（Content Analysis）

- 特長 （Holsti 1959; Krippendorff 1980; Riffe et al. 1988）
  - コミュニケーション「内容」
  - 長期間に渡って蓄積された「内容」を扱える
  - 対象の反応に影響されない（されにくい）
- 広い応用範囲 （Pool ed. 1959; Stone 1997; … ）
  - マス・コミュニケーション
    （新聞・ラジオ・テレビ）
  - 社会調査や心理学実験
  - その他： 民話、小説、会話
  - 実務的・商業的な局面：採用面接、市場調査

コーディング処理

- コーディングとは
  - サンプルをいくつかのカテゴリに分類
  - 分類することで、数え上げることを始めとする量的・統計的分析が可能に
  - カテゴリと分類基準（**コーディングルール**）
- 分析者の観点が反映される
  - 例えば「皺一つ無い制服を着た」人物
    - 「几帳面な人物」として数え上げるのか？
    - 「権威主義的な人物」として数え上げるのか？ Etc…
  - 理論仮説 / 問題意識の操作化 (Osgood et al. 1957)
  - アンケート調査で言えば質問文の作成

    上手く質問文を作成してこそ、回答者の意識・態度等を、引き出すことができる

量的方法と質的方法 （Lazarsfeld & Barton 1951）

- 分析の流れ
  - コーディングルールの作成（質）
  - 量的・統計的分析（量）
  - 図表の解釈 // 素データへの回帰（質）

  （必要に応じて 1. へ戻る）
- 互いに相補う関係

  「質的方法が洞察にあふれ、量的方法が仮説検証のための単に機械的なものだと決めつけるべきではない。この両者の関係は循環的なものである。すなわち、それぞれが新たな洞察をもたらし、それによって他方に資するものである」（Pool 1959: 192, 筆者訳）

テキストマイニングとの接合

- 提案：2段階での分析 （樋口 2004, 2005）
  - テキストマイニングによる探索的分析

    この段階では、恣意的なものとなりうる単語の取捨選択・編集（置換）は極力ひかえる
  - コーディングルールの作成

    ・探索的分析の結果を参考にしながら、

    ・目的に添って自由に単語を取捨選択・編集する

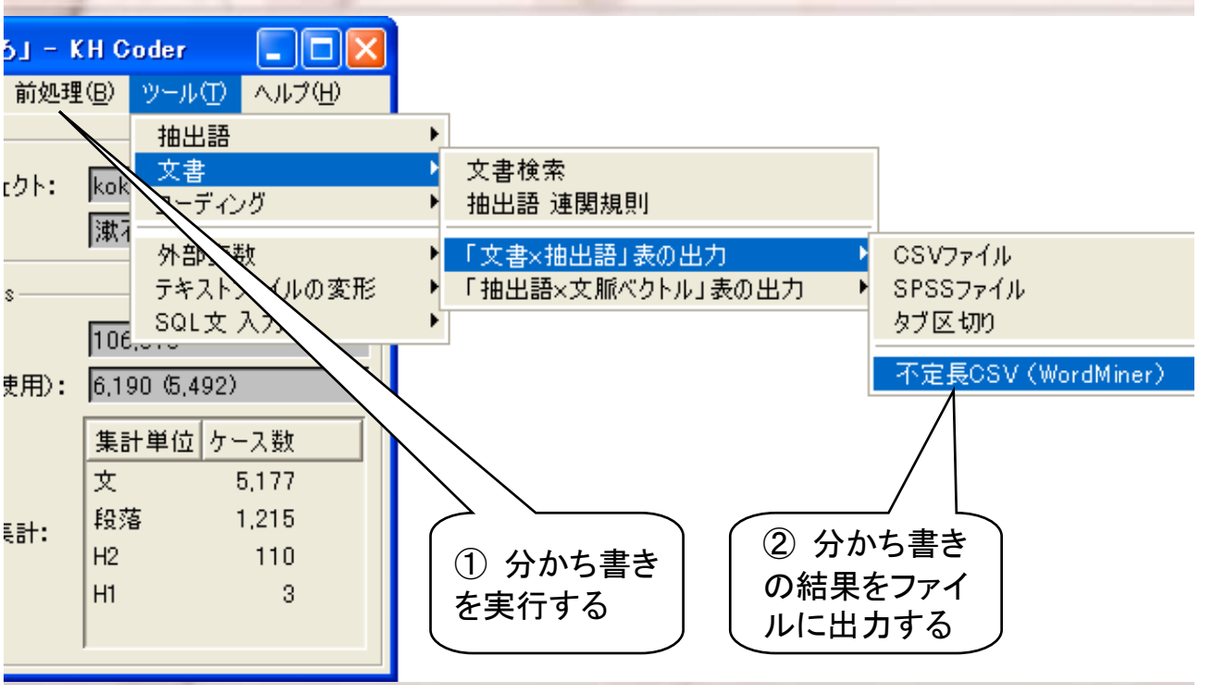
    （ここでは、「単語」だけにこだわらない）

    ・コーディングルールは開示する
- 「単語」を越えて：
  - 分析者の観点にもとづいた分類
  - 特に注目したい事柄を拾い出す
  - “disambiguation”：同じ言葉でも文脈によって意味が変わってしまうことに対応

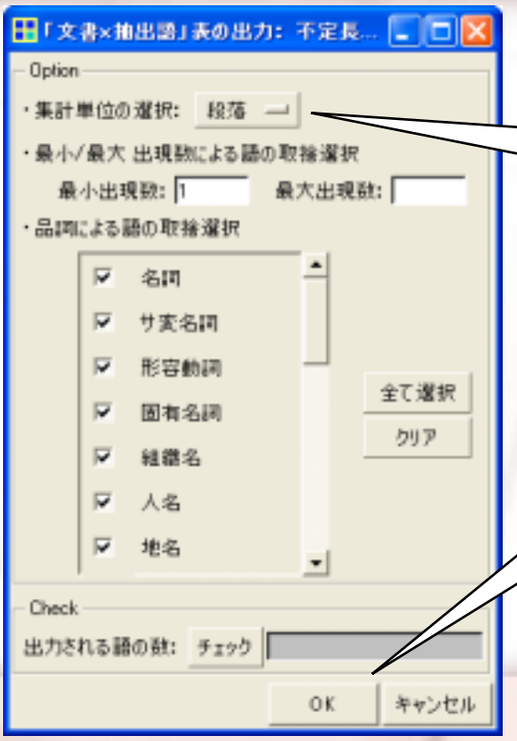
WordMinerとKH Coderの併用

- ソフト併用時の問題
  - 分かち書きの方式が異なると、同じデータから異なる単語が取り出される → 分析が煩雑に
  - 今回はKH Coder（茶筌）側にあわせることで、この問題に対処
    ※これによって、WordMiner上で品詞情報を利用できるようになる
- 分析・操作の流れ
  - KH Coderによる分かち書きの結果を、WordMiner上に読み込む ☆
  - WordMinerによる分析
  - KH Coderによるコーディング ☆
  - コーディング結果をWordMiner等で分析

分かち書き結果を出力（KH Coder）①



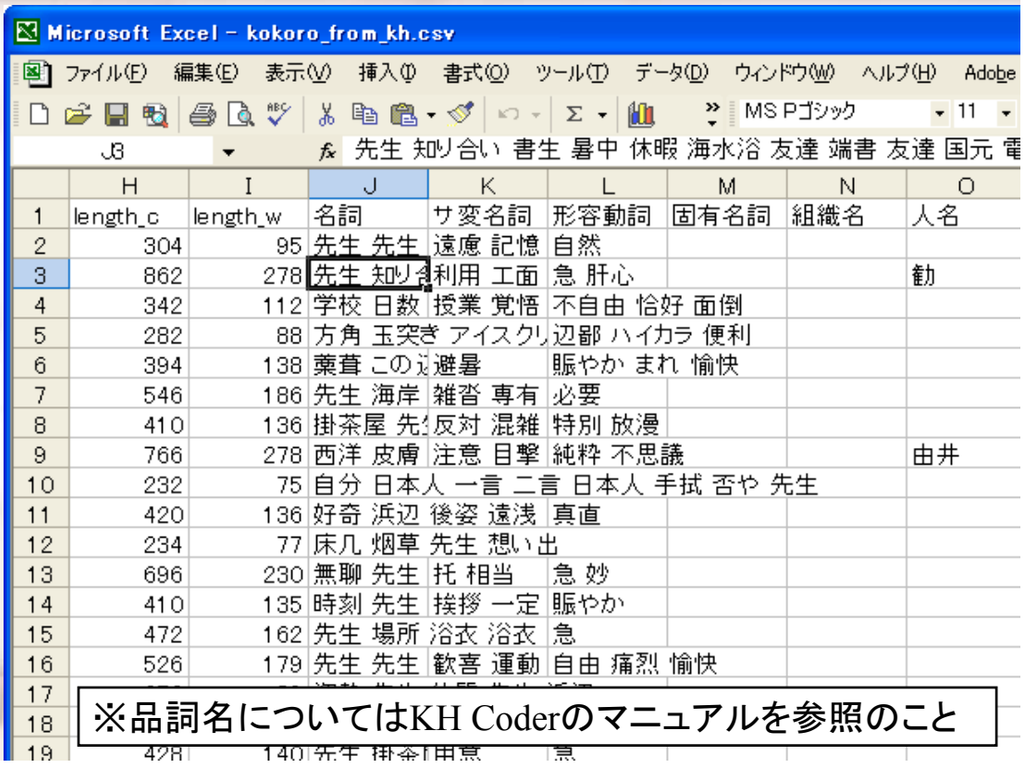
分かち書き結果を出力（KH Coder）②



①「段落」となっていることを確認する

②「OK」をクリック

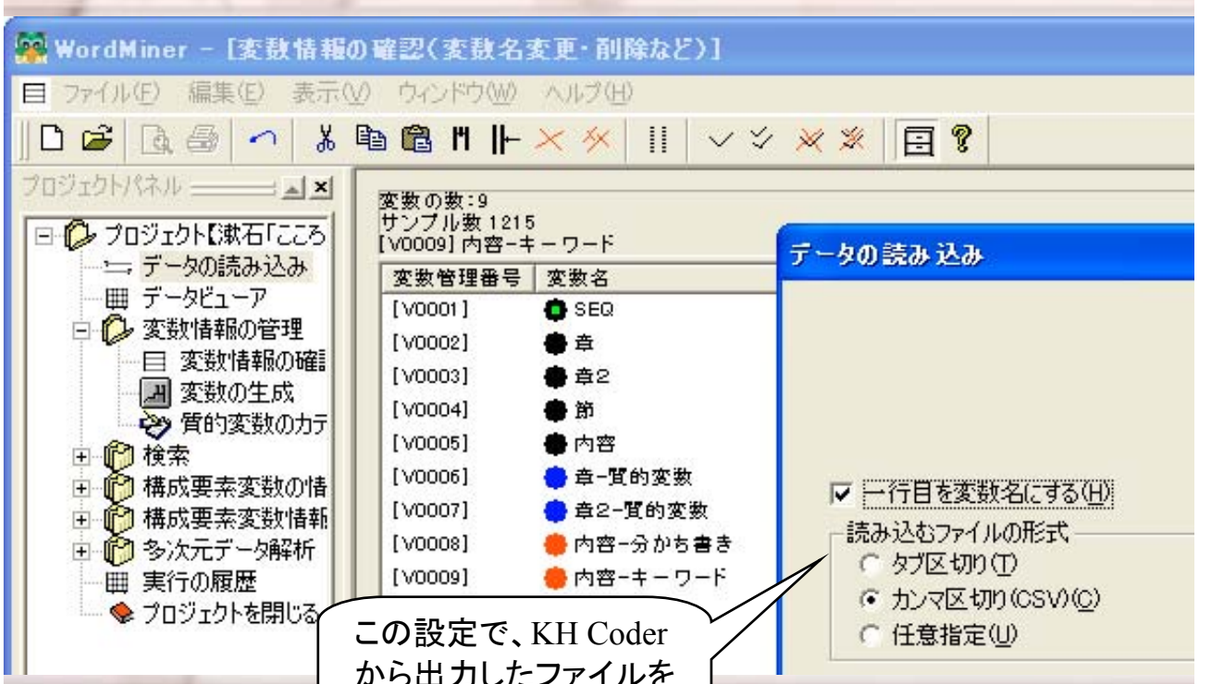
出力されるファイルの確認　※省略可



	H	I	J	K	L	M	N	O
1	length_c	length_w	名詞	サ変名詞	形容動詞	固有名詞	組織名	人名
2	304	95	先生 先生	遠慮 記憶	自然			
3	862	278	先生 知り合	利用 工面	急 肝心			勧
4	342	112	学校 日数	授業 覚悟	不自由 恰好 面倒			
5	282	88	方角 玉突き	アイスクリ	辺鄙 ハイカラ 便利			
6	394	138	薬草 この	避暑	賑やか まれ 愉快			
7	546	186	先生 海岸	雑沓 専有	必要			
8	410	136	掛茶屋 先	反対 混雑	特別 放漫			
9	766	278	西洋 皮膚	注意 目撃	純粋 不思議			由井
10	232	75	自分 日本人	一言 二言	日本人 手拭 否や 先生			
11	420	136	好奇 浜辺	後姿 遠浅	真直			
12	234	77	床几 烟草	先生 想い出				
13	696	230	無聊 先生	托 相当	急 妙			
14	410	135	時刻 先生	挨拶 一定	賑やか			
15	472	162	先生 場所	浴衣 浴衣	急			
16	526	179	先生 先生	歓喜 運動	自由 痛烈 愉快			
17								
18								
19			428	140	先生 掛茶屋	意		

※品詞名についてはKH Coderのマニュアルを参照のこと

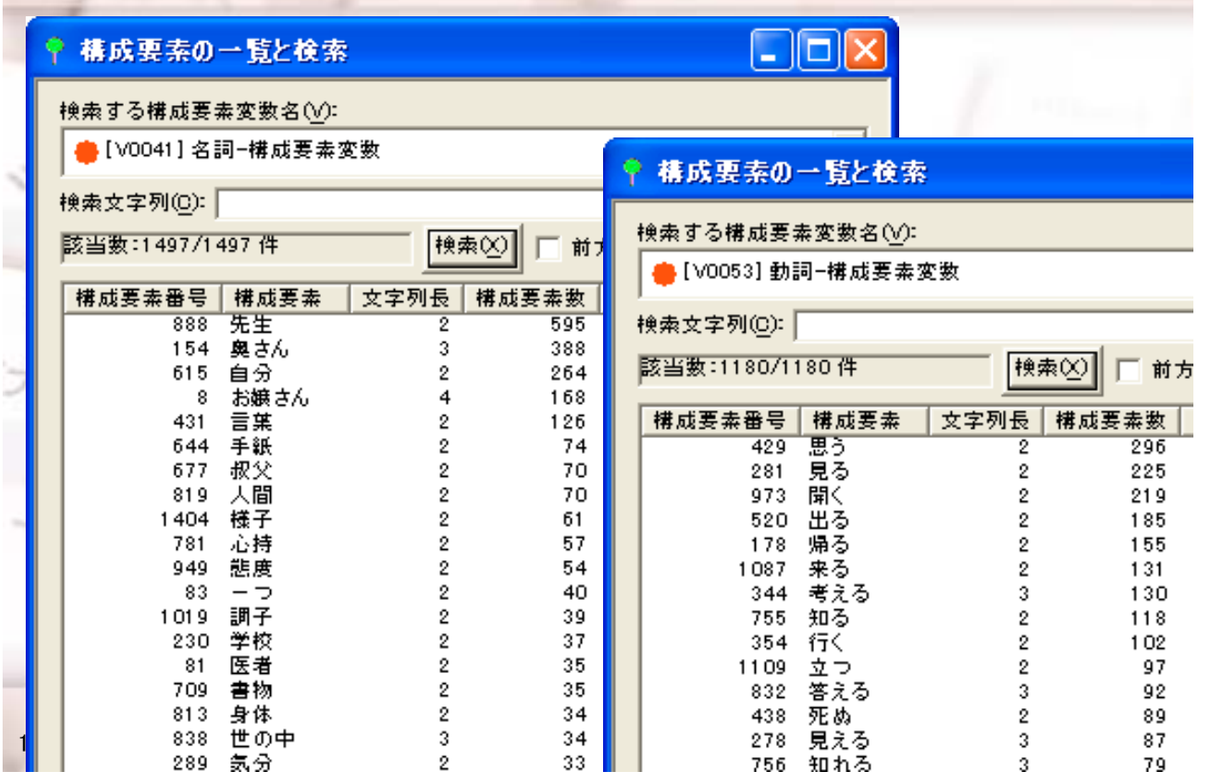
分かち書きの結果を読み込み



読み込んだ変数を「構成要素変数」に変換

「変数情報の管理」
→「変数の生成」
→「●構成要素変数を生成」
→「変数の種類を変更し、
新しい変数を生成」

生成した構成要素変数の確認　※省略可

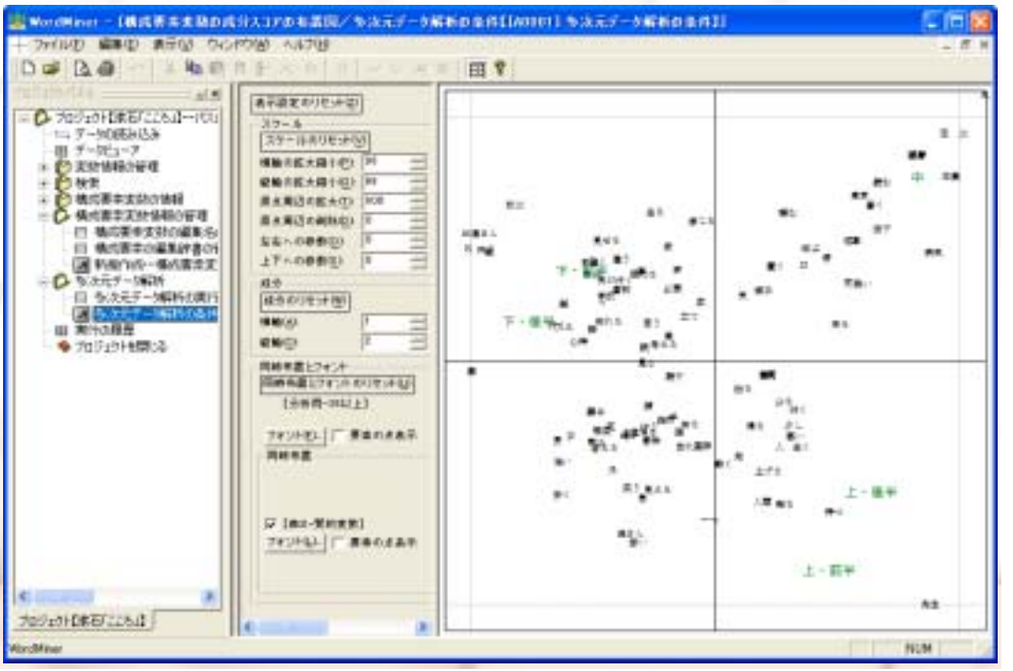


分析用変数の作成（品詞の選択）

「変数情報の管理」
→「変数の生成」
→「●構成要素変数を生成」
→「構成要素変数同士を併合し、新しい変数を作成」

WordMinerによる分析

– 以上の手順で、KH Coder（茶筌）の分かち書き結果を用いての分析が可能に



KH Coderによるコーディング①

- 指針：

  WordMinerによる探索的分析の結果を参考に、重要と思われる事柄を、条件指定（コーディングルール）によって拾い出していく

- 以下の条件を and / or / not / () で組み合わせる

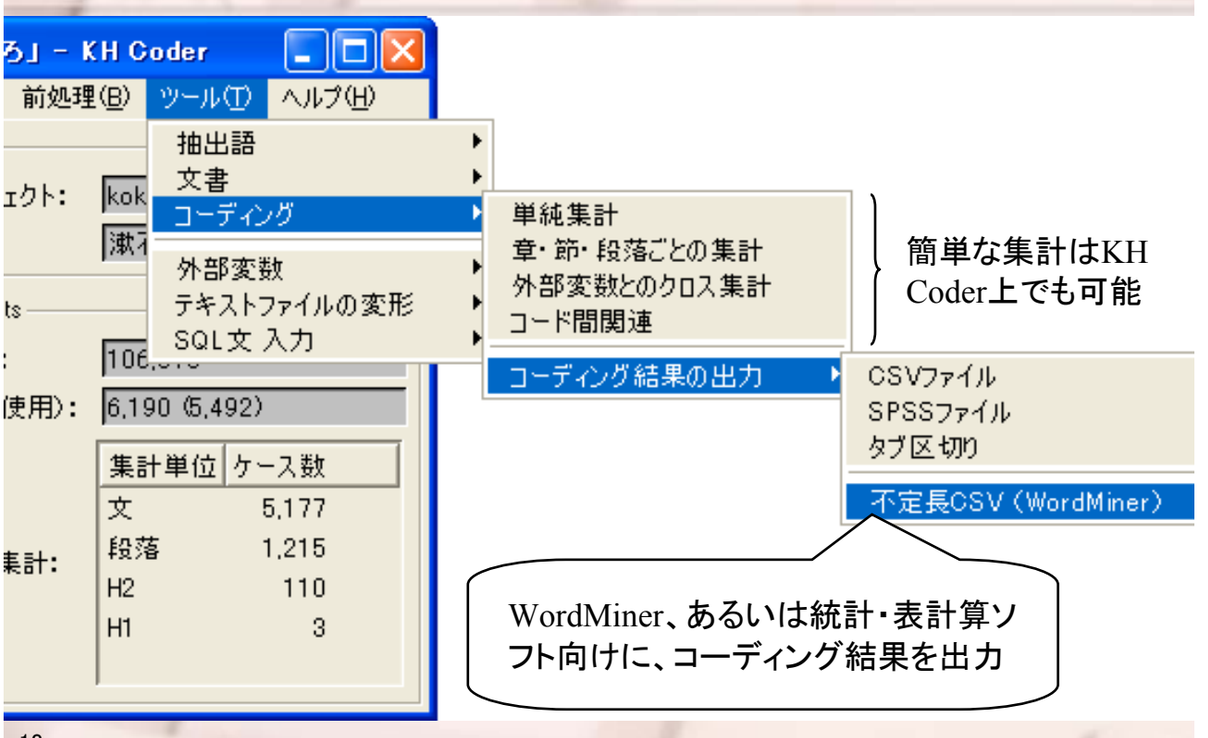
指定できる条件	条件の具体例
語の有無	「愛」という語が出現していれば
語のフレーズ	「卒業」と「論文」が連続して出現していれば
語の出現数	「愛」と「恋」があわせて3回以上出現していれば
外部変数	（データが自由回答の場合）女性の回答であれば
文書の長さ	（データが自由回答の場合）1語のみからなる回答であれば
文書の番号	（先頭から数えて）50番目以降100番目までの文書であれば
文字列	（自動抽出された語ではなく）「多ければ」という文字列が出現していれば

KH Coderによるコーディング②

- 単純なor接続
  - ＊告白　←コード名
  - 告白 or 打ち明ける or 自白 or 白状　←コードを与える条件
  - ※（機能としては）WordMinerによる置換と同じ
- Notの利用
  - ＊仕事
  - 仕事 or 会社 or (会議 and not 井戸端)
  - ※「井戸端での会議」は省かれる
- 変数の利用
  - ＊お嬢さん
  - お嬢さん or (奥さん and <>見出し1-->上__先生と私)
  - ＊奥さん
  - 奥さん and <>見出し1-->下__先生と遺書
  - ※「上」に登場する「奥さん」は、「＊お嬢さん」と見なす

➢ 必要なだけ、このような条件を作成する

コーディング結果の分析



文献

樋口耕一, 2004,「テキスト型データの計量的分析 ――2つのアプローチの峻別と統合」『理論と方法』19(1): 101–15.

――――, 2005,「計量テキスト分析の方法と実践」大阪大学大学院 人間科学研究科 平成16年度博士論文.

Holsti, O. R., 1969, *Content Analysis for the Social Science and Humanities,* Reading: Addison-Wesley.

Krippendorff, K., 1980, *Content Analysis: an Introduction to its Methodology,* London: Sage.(＝1989, 三上俊治ほか訳『メッセージ分析の技法 ――「内容分析」への招待』勁草書房.)

Lazarsfeld, P. F. & A. H. Barton, 1951, "Qualitative Measurement in the Social Sciences, Classification, Typologies, and Indices," D. Lerner & H. D. Lasswell eds., *The Policy Sciences: Recent Developments in Scope and Method,* Stanford: Stanford University Press, 180–8.

Osgood, C. E., G. J. Suci, & P. H. Tennenbaum, 1957, *The measurement of Meaning,* Urbana: University of Illinois Press.

Pool, I. d. S., 1959, "Trend in Content Analysis Today: A summary," *Trends in Content Analysis* 189–234.

Pool, I. d. S. ed., 1959, *Trends in Content Analysis,* Urbana: University of Illinois Press.

Popping, R., 2000, *Computer-assisted Text Analysis,* London: Sage.

Riffe, D., S. Lacy, & F. Fico, 1998, *Analyzing Media Messages: Using Quantitative Content Analysis in Research,* London: Lawrence Erlbaum Associates.

Stone, P. J., 1997, "Thematic Text Analysis: New Agendas for Analyzing Text Content," C. W. Roberts ed., *Text Analysis for the Social Sciences,* Mahwah: Lawrence Erlbaum, 35–54.