

NEPHIS - Nested String 索引法の 原理와 應用

川村敬一 著

高亨坤 訳

〈K. I. E. T. 綜合企劃分析室〉

1. 序 論

NEPHIS (NEsted Phrase Indexing System) 는 캐나다의 Western Ontario 大學의 Craven 이 考案한 新型의 順列索引이다. 出力方法은 現在, 印刷索引과 온라인에 의한 페이지(page) 表示, COM 피쉬(fiche) 등으로 되어 있다. NEPHIS의 概要는 1977년에¹⁾ 公表되었으나, 그후 實用化를 위한 評價와 온라인 情報處理의 새로운 地平을 開拓하기 위해 各種의 實驗에 使用되었다.^{2) 3) 4)}

Craven은 NEPHIS를 順列索引이라 말하나 그 原理는 從來의 所謂 自由順列(Free Permutation)法과는 相當히 다르다. 오히려 NEPHIS 索引法을 特徵 지우는 것은 스트링索引法과 構文上의 Embedding의 原理이다. 이러한 2가지 特徵을 同時에 갖춘 方式으로는 이외에도 PRECIS(Preserved Context Index System)⁵⁾ 와 ReI(Relational Indexing)⁶⁾ 이 있다. 또 分類를 基盤으로 한 類似方式으로는 印度의 POPSI⁷⁾ 가 있다. 이러한 스트링索引에서의 엔트리(Entry)는 用語의 리스트뿐만 아니라 룰(rule)에 따라서 編成되어 있다.

스트링索引法은 다른 여러 方式보다 精度가 높다고 하지만 그만큼 索引者의 負擔이나 作成費用도 많이 든다고 생각된다. 그러나 NEPHIS는 이러한 點을 考慮한 最新方式의 하나로서

모든 面에서 簡單히 設計되어 있다.

NEPHIS의 開發時 Craven이 考慮한 事項은 ①索引者의 코딩이 簡單하고, ②프로그래밍이 簡單하며, ③低廉한 價格으로 컴퓨터를 使用할 수 있으며, ④使用하기 쉽고 優秀한 出力形態일 것 등이다.

本稿에서는 먼저 NEPHIS의 코딩法과 알고리즘을 說明하고 다음으로 構文理論에서 導入된 Embedding의 概念과 스트링索引法을 究明한 후, NEPHIS의 原理와 實際에 대해서 記述하고자 한다.

2. NEPHIS의 作業順序

Nested란 같은 形態의 큰 것에 작은 것이 그 크기의 順序에 따라 차례 차례로 들어 가는 것을 의미하는 것이다. 예를 들면 5段 嵌套같은 것이다. NEPHIS의 基本方式은 이것과 恰似하며, 그 特徵은 무엇보다도 簡單함에 있는데 그것은 단지 4가지 코멘드記号(<, >, @, ?) 로써 索引作業을 할 수 있기 때문이다.

NEPHIS의 正式 主題索引方法은 먼저 論文의 內容分析을 하고 이것을 要約한 타이틀形式의 記述句를 作成하는데 이것은 NEPHIS의 先驅方式인 ASI⁸⁾의 入力句作成方法과 同一하다. 그 다음에는 記述句의 構文에 4가지 코멘드記号를 使用하여 코딩을 한다. 코딩된 記述句는 入力스트링(Input String)이라 불리우며, NEP-

HIS의 프로그램에 處理된다.

다음에는 예를 들어가며 코맨드記호와 알고리즘에 대해서 順次的으로 記述하고자 한다.

2.1 < , >

4가지 코맨드記号中 이 2가지는 變을 이루어 사용되는데 括弧는 Nested를 作成하기 위해 있다.

예컨대 Research Productivity of <Sleep Researchers>라는 스트링時, <Sleep Researchers>는 스트링 中의 Nested가 된다. 그래서 NEPHIS 프로그램은 다음과 같은 順列 엔트리를 作成한다.

○Research Productivity of Sleep Researchers

○Sleep Researchers. Research Productivity of —

前置詞에 의한 分節法은 ASI의 方法과 보다 恰似하다. 그러나 그것은 NEPHIS에서 더욱 複雜한 分節法이라기보다는 Nested作業이 可能한 것이다. 예컨대 < Teaching of <Information Science> at <University of <Strathclyde>>

위의 스트링時 <Strathclyde>는 <University of Strathclyde>의 또다른 Nested가 되어 있다. Nested는 한 變의 括弧로서 묶이므로 括弧의 數는 入力스트링에 있어서 左右가 똑같이 된다. 단, 제일 큰 入力스트링의 句에는 括弧를 하지 않는다. 또 形容詞+名詞, 名詞+前置詞+名詞의 構文에 Nested를 作成할 때는 恒常 後方の 名詞가 Nested가 된다.

엔트리는 다음 4가지 方法이 있다.

○Teaching of Information Science at University of Strathclyde

○Information Science. Teaching of — at University of Strathclyde

○University of Strathclyde. Teaching of Information Science at —

○Strathclyde. University of —. Teaching of Information Science at —

프로그램은 各己의 語句가 Nested되었는지의 與否를 確認하고, Nested된 경우는 피어리어드

와 스페이스가 附与되어 Nested를 包含하는 句의 나머지 部分이 添加된다. 그리고 Nested가 抽出된 部分은 뎃쉬로 換置된다. 이러한 操作이 모든 Nested에 대해서 反復된 다음 入力時의 코맨드記号는 消滅되고, 索引語의 ABC順으로 配列된다.

2.2 @

Nested의 앞에 @가 附与된 때는 거기에서 順列이 始作되지는 않는다. 이 記号는 Analysis, Effect, Study 등의 用語에 附与되고 KWIC의 不用語(Stopword)의 概念에 該當된다. 예컨대 Equations for <@ Design of <Retrieval Systems>>의 스트링에서 다음의 엔트리가 생겨난다.

○Equations for Design of Retrieval Systems

○Retrieval Systems. Design of —. Equations for —

2.3 ?

이 記号는 順列에 例外를 만든다. ? 다음의 文字列은 限定된 狀況에서만 엔트리에 出現한다. Craven에 의하면 判讀力(readability)이나 明瞭度(intelligibility)를 增加시키기 위함이며 ?에는 2가지 用法이 있다.

2.3.1 前方判讀連結語태그

前方判讀連結語(forward-reading connective : 以下 frc라 略함)는 ?……<의 形으로 表示되며, 태그는 Nested의 始初를 表示한다. frc는 그것을 包含하는 큰 句로부터 順列이 始作될 때, 有效한 連結語가 되나 frc의 다음 Nested로부터 始作될 때는 無效한 스트링(null string)으로서 除去된다. 또한 Nested의 抽出을 알리는 뎃쉬도 省略된다.

frc태그를 使用하는 理由는 예를 들어 Books. Collection of — 라는 엔트리가 있을 때 of는 Books를 目的語로 하는 後方指示前置詞(backward-pointing preposition)이다. 그러나 NEPHIS에 있어서 이러한 種類의 前置詞는 意味를 理解하기 위해 반드시 必要한 것은 아니다.

그러므로 of를 除去하는 것이 判讀하기 容易하다고 Craven을 말한다. 예컨대 <Forecasts ? on <Publication ? in <Science>>의 스트링時 frc인 on과 in은 各已 前方에 目的語를 가질 때 만 엔트리에 出現한다.

- Forecasts on Publication in Science
- Publication in Science. Forecasts.
- Science. Publication. Forecasts

만약 위의 入力스트링에 다시 語句가 追加되었다고 한다면

- Forecasts ? on <Publication ? in <Science>> ? in <Japan>

이때 앞의 2번째 엔트리는 다음과 같이 된다.

- Publication in Science. Forecasts in Japan

Publication in Science가 先導語句로서 抽出되었으므로 Forecasts는 그대로 in Japan에 따른다. Forecasts에 따르는 前置詞는 當然한 일로서 그 目的語에 의해서 on과 in으로 區別된다.

2.3.2 後方判讀連結語태그

後方判讀連結語(backward-reading connective : 以下 brc라 略함)는 ?……>의 形으로 表示되고, 태그는 Nested의 끝을 指示하고 brc를 無效로 한다. 그러나 brc는 그것을 包含하는 Nested에서 順列이 始作되어 後方으로 되돌아갈 때는 有效한 連結語가 된다. 또 brc는 피어리어드와 스페이스의 代用도 된다. 그리고 많은 경우에 and나 or로써 表現되는 同位關係에 frc와 함께 使用된다. 예컨대 Information Science ? & <Operations Research ? &>의 스트링時 앞의 &는 frc이고 뒤의 &는 brc이다. 最初의 엔트리에서는 뒤의 &가 除去되고 다음의 엔트리에서는 앞의 &가 除去된다.

- Information Science & Operations Research.

- Operations Research & Information Science.

brc태그는 初歩者에게는 어려울 때도 있으나, 이것을 마스터하며는 다음과 같은 自然語表現을 코딩하는데 有用하다.

- Title ? <① Compared with ? <Abstracts ?>?>
- Titles Campared with Abstracts
- Abstracts Compared with Titles

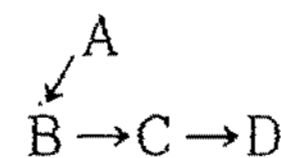
3. Embedding의 概念

말로 하는 말이든, 쓰는 말이든, 우리들이 日常 使用하는 言語가 때로는 意味를 옮겨 傳達하지 못하는 原因의 하나는 그것이 線形配列(linear order)의 制約을 받고 있기 때문이다. 이러한 制約을 補完하기 위해 우리들은 冂形의 手段을 갖고 있다. Embedding의 概念이 言語學에 導入된 것도 오래인데 여기에서도 構文을 說明하기 위해 冂形이 活用되어 왔다.

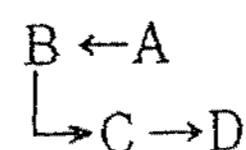
Hutchins⁹⁾는 索引言語에 있어서의 Embedding의 典型을 PRECIS나 ReI에서 찾고 있다. 複合主題를 表示할 때 索引語의 配列法에는 5가지의 基本形態가 있는데, 그것을 그림 1로 說明해 보았다. 즉, 主題를 表現하는 索引語 原來의 並列이 ABCD일때 派生하는 索引語의 並列을 5가지 形態로써 簡單하게 表示한 것이다.

3.1 PRECIS의 Embedding

그림 1의 5는 Austin의 PRECIS를 모델로하고 있다. 파시트分析에 따라서 役割演算子(role operator)를 附与한 索引語는 入力스트링에 있어서 階層上에서는 거의 大部分 概念의 大에서 小로 되고, 또 構文上에서는 受動態에 恰似한 並列方法으로 된다. 즉, A→B→C→D이다. 그러나 여기서는 A 以外에 즉, B가 先導語(Lead)가 되었을 경우 스트링은 크게 3個의 部分, 다시 말해 B와 A와 CD로 分割될 수 있다.



PRECIS에서는 2行式의 表示法을 취하고 있는데, 先導語와 같은 行(Qualifier의 位置)에는 그것을 限定하는 上位概念을, 그리고 아래 行(Display의 位置)에는 下位概念을 配置한다.



1.permutation	2.cycling	3.rotation	4.truncation	5.embedding
ABCD	ABCD	A BCD	DCBA	ABCD
ABDC	BCDA	A B CD	CBA	B(A)CD
ACBD	CDAB	AB C D	BA	C(BA)D
ACDB	DABC	ABC D	A	D(CBA)
ADBC				
ADCB				
BACD				
BADC				
etc.				

(4, 5: A>B>C>D)

그림 1. 索引語의 表示法

이것은 索引語를 한 方向으로 循環시킨 循環시킨 配置法으로 看做할 수 있으나, 同時に 先導語를 分岐点으로 하여 2次元의 意味表示를 可能케 한다. 즉, 첫째 行에서는 그림 1의 4처럼 切頭方式의 代表인 Chain Index와 같은 逆配列에 의한 文脈單位를 表示하고, 두번째 行에서는 入力스트링과 같은 配列에 의한 文脈單位를 表示한다. 만약 이것을 從來의 첫째 行의 線形 配列로 하면 Embedding이 行해져서 다음과 같이 될 것이다.

$$B(\leftarrow A) \rightarrow C \rightarrow D$$

PRECIS의 2行式 表示法은 圖形的 痕跡을 남기고 있으나 配列順序는 Embedding의 原理에 따라서 各 엔트리의 Lead, Qualifier, Display의 ABC順으로 되어 있다.

3.2. Rel의 Embedding

Farradane의 Rel에서도 똑같은 Embedding이 行하여진다. 關係演算子(link operator)로 아래의 2가지 意味構造(diagram)가 表示되어 있을 때 과연 어느 쪽이 먼저 配列되어야 할 것인지의 問題를 解決하기 위해 Embedding이 行하여 진다. 즉, 線形으로 變換한다.

$$\begin{array}{l} \textcircled{1} A / : B / ; E \\ \quad / \\ \quad C \end{array} \qquad \textcircled{2} A / : B / ; D$$

①은 $A / : B [/ - C] / ; E$ 와 같이 되어 B와 C와의 關係를 破壞치 않고 主題를 表示할 수

있다. 이 경우의 配列은 ①이 먼저가 되나 ①이나 ②모두가 A 以外の 概念을 先導語로 할 때는 Embedding에 의하여 主題의 完全表示를 행한다. 예컨대 ①은 다음과 같이 된다.

$$\begin{array}{l} \circ B [: / A] [/ - C] / ; E \\ \circ C - / B [: / A] / ; E \\ \circ E ; / B [/ - C] : / A \end{array}$$

3.3 NEPHIS의 Embedding

크게 말해서 NEPHIS의 作業順序는 變形生成文法(transformational-generative grammar)의 操作의 一部로서 말할 수 있을지도 모른다. 이 文法理論에서는 日常言語의 表層構造(surface structure)는 意味를 記述하는데 不充分하며, 文法的으로 一般성이 높은 深層構造(Deep structure)라는 概念이 導入되어 있다. 그리고 Embedding Sentence를 中心例로 하여 表層構造上的 2種 以上の 同義文 중에는 一聯의 置換·消去 등의 變形을 거쳐서 同一의 深層構造에서 生成되고 있음을 例証한다. 우리들에게 잘 알려진 Embedding Sentence는 英語의 關係詞節이나 Chomsky에 따라서 한 예를 들어 보기로 한다.

"a wise man is honest"라는 文의 表層構造는 主部 "a wise man"과 述部 "is honest"로 分解될 것이다. 그러나 深層構造는 좀 다르다. 이것은 유럽語에 있어서 S + V + C 構造의 文은

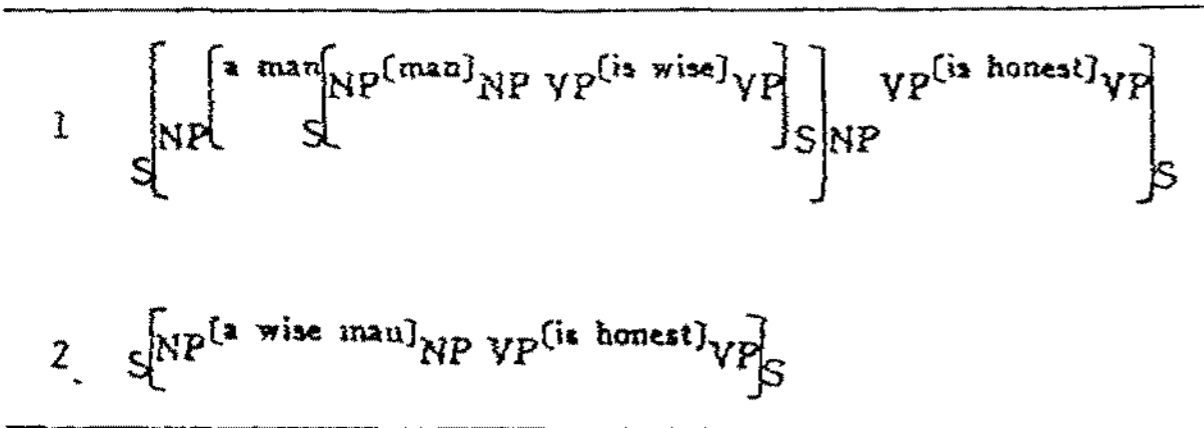


그림 2. 名稱을 갖은 括弧 附与하기

命題(propositon)를 根底에 形成하고 있기 때문이다. 그래서 表層構造의 主部를 構成하는 複合概念에서 다시 主語 "man"과 述語 "be wise" 로써 表示되는 根底命題를 찾아낼 수가 있다. 이와 같은 命題의 所在는 句構造(Phrase Structure)에 名稱을 갖은 括弧 (labeled bracketing) 附与함으로써 明白히 된다. 上記 文의 深層構造와 表層構造를 括弧를 주어 表示하면 各己 그림 2의 1과 2와 같이 된다.

1과 2의 어느 쪽이나 NP(名詞句)와 VP(動詞句)로써 構成되어 있으나 深層構造는 意味를 限定하고 表層構造는 音聲表示를 明白히 하고 있다. 즉, 1은 "a man"을 다시 限定한다. 이 事實을 說明하기 위해 그림 2의 括弧附与를 土臺로 構文上의 句構造의 階層을 나타낸 것이 그림 3이다. 여기에 이르러 1에서 Embedding으로 看做된다.

이 예의 深層構造에서 表層構造로의 變形操作은 Chomsky¹⁰⁾의 論述을 参照하기 바란다. "a wise man is honest"와 "a man who is wise is honest"는 同一한 深層構造에서 派生한 文이지만 後者の 쪽이 그림 2의 1의 根底命題를 表示하는 文에 가깝다.

上述의 構文分析中에서 NEPHIS와 關係가 있는 것은 括弧附与하기 이다. NEPHIS의 括

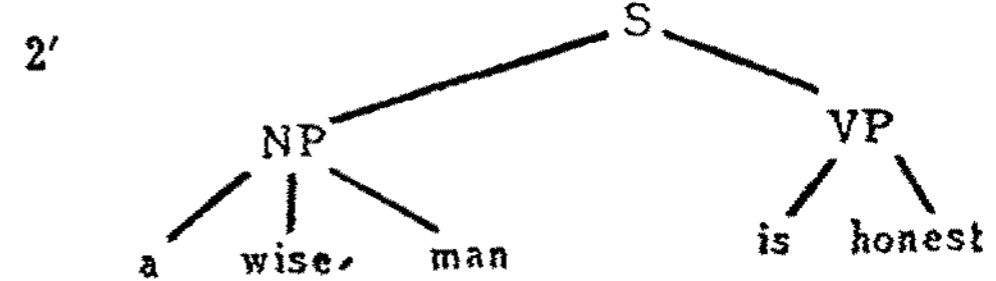
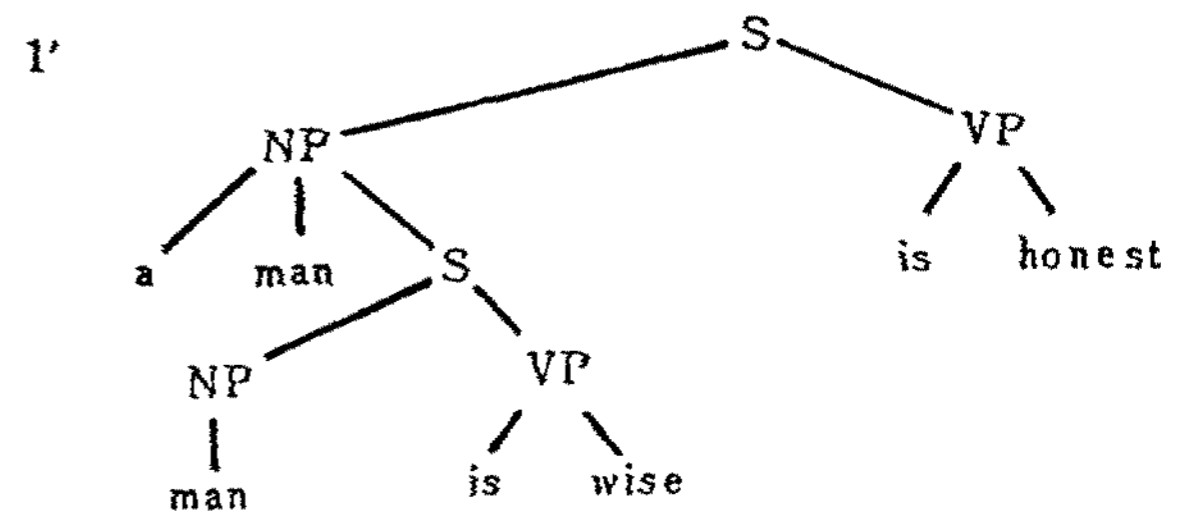


그림 3. 句構造의 階層

弧附与(Nested 作業)는 記述句에 숨겨진 Embedding의 事實을 明白히 한다. 이것은 線形으로 明示할 수 없는 名詞句의 關係를 한 번 圖形的으로 表示하고자 하는 作業이기도 하다. 그림 4는 Nested의 階層關係와 統語關係의 한 예이다.

많은 경우 下位의 Nested는 上位의 그것을 限定한다. 또 統語關係는 順列規則을 정한다.

한편 Craven은 NEPHIS의 시뮬레이션 시스템을 開發中이다. 마이콤으로도 運營이 可能한 이 시스템은 먼저 가지고 있는 入力 스트링에서 프로그램이 自動적으로 Nested의 削除·抽出·置換·挿入 등의 操作을 행하고 擬似스트링(pseudo strings)을 作成한다. 그 다음 入力 스트링과 擬似스트링에서 擬似엔트리(pseudo entries)가 作成되어 擬似索引이 CRT스크란에 表示되는 裝置이다. 機械可讀의 디소러스에 連結될 수 있으면 더욱 強力해지는데, 어찌되었든 利用者는 最初의 主題表明으로 類似의 主題를 推論할 수 있다.

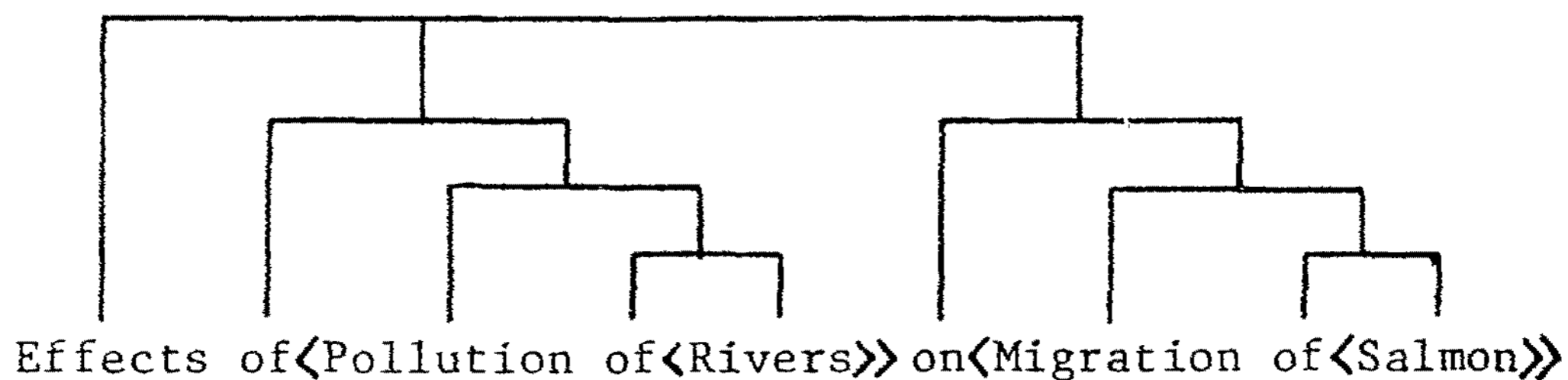


그림 4. Nested의 階層關係와 統語關係

4. 스트링 索引法(String Indexing)

4.1 概念의 限定과 完全한 主題表示

스트링 索引方式의 順列規則은 1個의 言語體系에서의 書換規則에 対応시킬 수 있을지 모른다. 翻譯하기 어려운 用語이기는 하지만 Svenonius¹¹⁾ 등에 따르면 어느 複合主題를 構成하는 各 概念이 孤立的은 아니고 相互間에 特定의 關係를 保持하는 것으로 取扱되는 方法이다. 이 關係는 이어지는 스트링, 즉 演算子와 그것을 附与한 索引語의 並列에 의해서 하나의 主題로서 設定된다. 各索引語의 關係는 入力스트링의 特定 文脈中에서 限定되므로 索引語의 順列法에는 스스로 文法的인 制限이 가해진다. 또한 어떤 索引語가 先導語가 되든지 엔트리는 原來의 主題와 同一한 意味를 完全히 表示할 수 있다.

4.2 스트링索引法의 起源

이 索引法은 前章에서 본 것과 같은 Embedding의 原理에 따르는 方式으로 代表되나 다른 性格을 갖는 것도 包含하여 많은 적들간에 言語學의 諸原則에 따른다. Svenonius¹¹⁾ 등은 스트링索引法의 到來를 予告한 것은 랑가나단(Ranganathan)의 콜론分類法이었다고 한다. 그의 手法의 根幹은 分析合成方式(analytico-synthetic processing)으로서 有名한데 이것은 파시트(facet)에 의한 意味分析과 파시트公式에 따르는 統語的 合成方式으로 이루어진다. 또한 그의 Rounds의 概念은 言語學에 있어서 Embedding의 概念과 明白히 対応하는 것이다.

5. PRECIS, ReI, NEPHIS의 例

그림 5는 一般的으로 複雜한 主題를 PRECIS, ReI, NEPHIS의 各 코딩法과 그 出力形態를 表示한 것이다.

PRECIS가 26개의 役割演算子와 9개의 關係演算子를 使用하는 반면, NEPHIS에서는 4개의 코멘트記호를 記憶하는 것만으로도 充分하다. 또 相互参照(See, See also)의 엔트리도 入力

PRECIS *Effects of river pollution on the migration of salmon*

(x) (1) salmon*
(y) (2) migration*
(y) (2) effects of river pollution
(x) (1) rivers*
(y) (2) pollution*
(y) (2) effects on migration of salmon

Migration. Salmon
Effects of river pollution
Pollution. Rivers
Effects on migration of salmon
Rivers
Pollution. Effects on migration of salmon
Salmon
Migration. Effects of river pollution

RELATIONAL INDEXING

river
/
salmon/-pollution
/*
migrating
Migrating
of salmon. Affected by pollution of river.
Pollution
of river effect on salmon migrating.
River
pollution effect on salmon migrating.
Salmon
migrating affected by pollution of river.

NEPHIS

@Effects? of (Pollution? of (Rivers)? on (Migration? of (Salmon))
Migration of Salmon. Effects of Pollution of Rivers
Pollution of Rivers. Effects on Migration of Salmon
Rivers. Pollution. Effects on Migration of Salmon
Salmon. Migration. Effects of Pollution of Rivers

그림 5. PRECIS, ReI, NEPHIS의 例

스트링에 2가지 記号(=, ==)의 한 쪽이 追加되는 것만으로도 順列的으로 作成할 수 있다.¹²⁾

그러나 모든 同一主題가 그림 5와 같은 同一 엔트리가 된다고 限定하지는 않는다. 이것은 단지 主題分析法이나 記述句의 作成法에 起因할 뿐이다. 적어도 NEPHIS에서는 ① PRECIS와 같은 파시트分析이나 ReI와 같은 認識 또는 思考의 心理分析(및 파시트分析)은 施行되지 않는다. ② 또 入力스트링에 있어서 標準的인 索引語의 並列方法(standard citation order)도 規定되어 있지 않다. 記述句의 構文에 直接코딩함으로써 意味와 統語의 分析은 거의 同時에 行하여진다. 여기에 NEPHIS의 簡單성과 限界성이 共存하는 것으로 생각된다.

自然語的 記述句의 真正한 統語關係는 무엇인

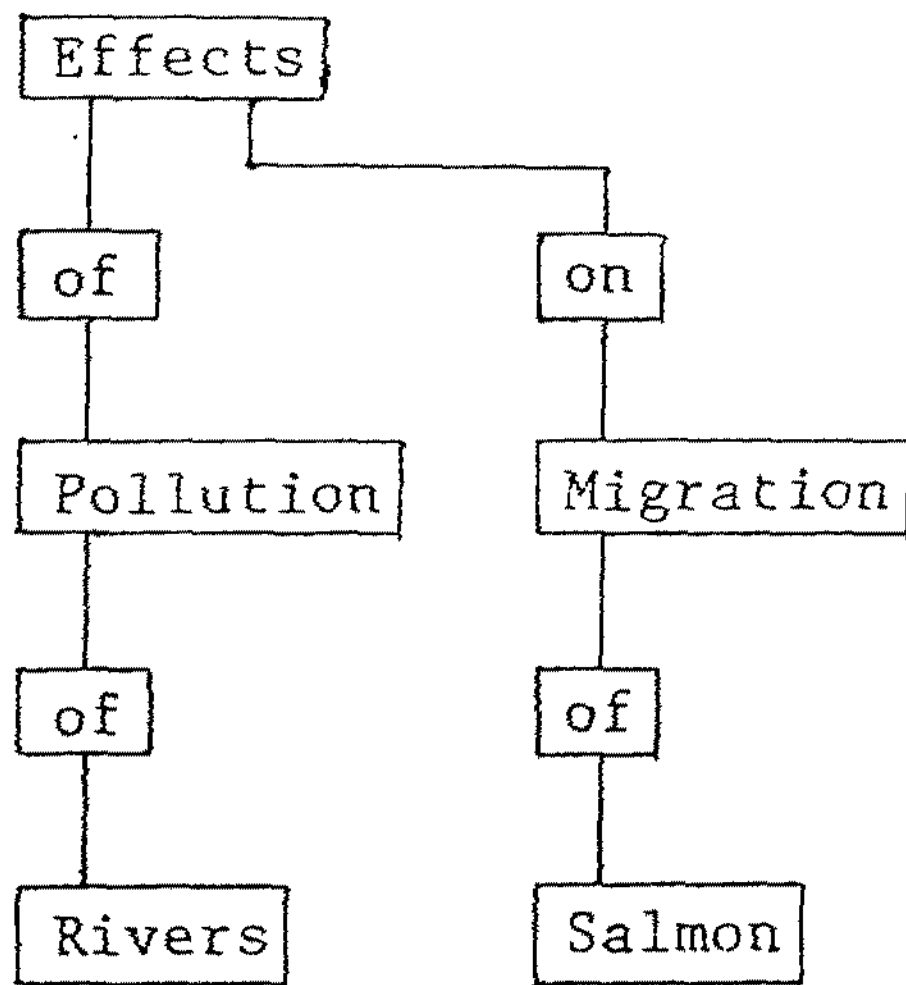


그림 6. 概念의 링크

가의 方法으로 明確히 되지 않으면 안된다. N-EPHIS의 Nested作業이 이를 위한 方法이라는 것은 이미 言及한 바 있다.

그림 6은 Craven이 概念링크(concept links)라 呼稱하는 것으로서 그림 4의 Nested의 統語關係를 簡潔하게 表示하고 있다. 링크의 基本的인 形態를 類別하는 것도 重要한 課題이다. '어쨌든간에 링크에 따르며는 그림 5와 같은 適切한 엔트리를 얻을 수 있다.

또 上記 3가지 方式의 比較를 위한 豊富한 예는 Craven과 Farradane이 所属한 Western Ontario 大学刊行의 매뉴얼에 記載되어 있다. 이 매뉴얼은 3冊1組로 되어 있고, J. Am. Soc. Inf. Sci. (美国情報科学会誌) 24~25卷(1973~1974)에 収録된 91件의 同一論文의 抄錄을 基礎로 各考案者가 独自の인 方式으로 만든 것이다. ^{13) 14) 15)}

6. NEPHIS 索引法の 要点과 補助用具

6.1 要点

Craven은 "NEPHIS는 簡單하나 결코 事務的인 作業은 아니다"라고 말했는데, 특히 Nested作業은 重要하고 誤差는 엔트리에 重大한 影響을 미친다. 이것은 NEPHIS가 ASI를 先驅方式으로 하면서도 ASI와 같은 純然한 統語시

1	Destruction? of<Bacteria>? by<Dyestuffs> Destruction of Bacteria by Dyestuffs Bacteria. Destruction by Dyestuffs Dyestuffs. Destruction of Bacteria
2	Destruction? of<Bacteria? by<Dyestuffs>> Destruction of Bacteria by Dyestuffs Bacteria by Dyestuffs. Destruction Dyestuffs. Bacteria. Destruction

그림 7. 코팅의 正誤例

1	Management? of<Railways? in<France>> Management of Railways in France Railways in France. Management France. Railways. Management
2	Management? of<Railways>? in<France> Management of Railways in France Railways. Management in France France. Management of Railways

그림 8. 2種의 코팅法

스팀이 아니고, 약간의 意味 分析이 따르기 때문이다. 예컨대 그림 7에서는 1이 옳고, 2는 틀리는 것이다. 엔트리를 보면 알 수 있는데 2의 두번째에서는 Bacteria와 Dyestuffs에 作用의 中間項이 必要하고, 세번째에서는 "Destruction of Dyestuffs by Bacteria"와 区分할 수 없기 때문이다.

索引作業의 一貫性도 시스템의 重要한 課題이다. 그림 8은 場所의 意味를 取扱한 예이다. 이 경우는 1이 옳으나 内容에 따라서는 2도 있을 수 있으므로 索引者에게는 正確한 指針이 必要하다. 그러나 周到한 PRECIS에서는 場所의 意味를 정하는 役割演算子가 研究의 觀點에서 区别하고 있으므로 틀림없이 다음과 같은 1의 配列이 많은 엔트리가 된다.

- Management. Railways. France.
- Railways. France.
Management
- France
Railways. Management.

6.2 매뉴얼

誤差에 대한 注意나 一貫성을 위한 指針으로서 먼저 매뉴얼이 必要하다. 前章에서 NEPHIS의 매뉴얼을 紹介하였지만 複雑한 主題를 豊富하게 表示하는데 不過하므로 不充分하다. Batty도 本書의 價值를 定하는데 있어서 다른 2卷과 마찬가지로 그것을 매뉴얼이라기보다는 데몬스트레이션이라고 보고 있다.^{16) 17)}

ReI는 긴 歷史를 갖고 있으며, PRECIS는 1974년에 大部分의 매뉴얼이 만들어졌다. 그리고 이번에는 英國에서 NEPHIS教授用의 패키지(워크북과 매뉴얼)가 刊行되었다. 作成의 指揮를 担当했던 Keen에 의하면 이것은 마이콤을 사용한 實驗教授의 報告書(British Library R & D Reports)로서, Commodore PET 마이콤用으로 作成된 BASIC프로그램은 KWAC도 出力可能하게 되어 있다. NEPHIS코딩은 KWAC의 先導語도 同時에 選定하고 있기 때문이다. 워크북은 文字 그대로 練習問題로 되어 있고, 매뉴얼에는 教授法과 프로그램 등이 詳細하게 記述되어 있다.

6.3 図形에 의한 診斷用具

機械的인 配列을 하는 KWIC는 別途로 하고 一般的으로 規則에 따라서 엔트리가 編成되는 computer-formatted方式에서는, 索引者는 複雑한 主題에 대하여 出力形態를 念頭に 두고서 코딩할 必要가 있다. 이미 ASI를 採用하고 있는 Chemical Abstracts에서는 엔트리의 品質管理를 위해 CRT스크린을 使用하는 온라인 編輯法을 實施하고 있다.¹⁸⁾

Craven도 NEPHIS를 위해 마이콤을 使用한 強力한 診斷用具(diagnostic tool)를 開發하였다. 그리고 코딩을 終了했을 경우 不均衡한 括弧附與가 체크되는데, 그림 6은 概念의 링크가 스크린에 表示되는 裝置를 나타낸 것이다. 索引者는 図形에 의해서 統語關係를 確認한 後 最終 데이터를 中央의 컴퓨터에 蓄積할 수 있다.

7. KWOC vs NEPHIS

NEPHIS의 實用化는 科学技術文献의 타이틀에 適用되는 것으로서 始作되었으며, 対象資料는 캐나다政府의 에너지·鉦業·資源省에서 刊行된 1966~78年 까지의 約 3,000件의 技術리포트類이다.¹⁹⁾ 實驗을 겸한 이 試圖에 대하여 Craven 등은 National Research Council of Canada와 Natural Sciences & Engineering Research Council에서 助成金を 얻었다.

이 實驗의 目的은 ① 타이틀을 索引語로 하는 KWOC과의 比較, ② 이미 말한 正式의 主題分析을 施行한 NEPHIS와의 比較이다. 그리고 그는 「NEPHIS가 KWOC보다도 檢索時間이 빠를 것」이라고 予測하였다. ①에 대해 簡單히 말해 두고자 한다.

公平을 기하기 위해 같은 타이틀에서 KWOC과 NEPHIS를 作成하고, 質問도 타이틀에서 시뮬레이션되어 作成되었다. 이 比較에서 兩者의 相異點은 索引語의 配列法 뿐이다. 그 結果는 다음과 같다.

① 코딩 : KWOC에서는 不必要하지만 NEPHIS는 그것을 위해 時間과 人件費를 要한다.

② 코딩의 誤差 : NEPHIS入力스트링의 4.8%는 프로그램이 處理되지 않았다. 修正을 위해 2번째의 作業을 행하고, 1.2%로 減少시켰다. 이외에 發見이 容易치 않은 Nested의 誤差가 있었다. 이것들은 全엔트리의 5~8%에 影響을 미쳤다(단, 図形에 의한 診斷用具는 이 實驗에서는 使用하지 않았다).

③ 페이지數 : 兩者 모두 冊子体로 約 250페이지라는 報告이다.

④ 檢索時間 : 1회의 檢索에 NEPHIS가 KWOC보다 빨랐다(NEPHIS는 39.9秒, KWOC은 47.4秒).

⑤ 브라우징機能(browsability) : NEPHIS가 특히 優秀하다.

이 結果 NEPHIS에 要求되는 負擔은 ①과 ②이다. 그러나 그 簡單함을 생각하면 ④나 ⑤의 어느 쪽도 負擔을 相殺하는데 充分하다고 Craven은 말하고 있다. 結局 当初의 予測은 옳

았으나 理由는 兩者가 모두 文脈을 保持하는 方式이면서도 NEPHIS에서는 KWOC와 같이 有節句를 逐語的(sequential)으로 읽을 必要가 없기 때문이다. 타이틀의 경우뿐만 아니고 이러한 特徵은 NEPHIS나 PRECIS가 KWOC, KWIC, ASI 등의 엔트리形態보다 優秀한 点이다. 그는 NEPHIS의 타이틀 코딩에는 多少의 追加經費가 必要하나 KWOC이나 KWIC보다도 優秀한 索引을 할 수 있다고 하고 있다. 이 實驗의 副産物로서 타이틀 코딩用的 매뉴얼 改訂版도 刊行되었다.²⁰⁾

8. 結 語

最新의 索引方式인 NEPHIS의 原理와 応用 시스템에 대해서 지금까지 말해왔다. 그 原理는 簡單하기는 하나 英語構造의 洞察에서 考案된 優秀한 것으로 評價되어도 좋다. 또 마이컴을 使用한 応用시스템은 온라인 情報處理의 技術을 言語와 心理의 面에서 發展시키는 것으로서 注目해야 할 것이다.

〈参 考 文 献〉

- 1) Craven, T. C. : NEPHIS : a Nested-Phrase Indexing System. J. Am. Soc. Inf. Sci. 28(2), 107-114(1977)
- 2) Craven, T. C. : Microcomputer-Generated Graphic Displays as an Aid in String Indexing. J. Am. Soc. Inf. Sci. 31 (1), 123-124 (1980)
- 3) Craven, T. C. : Simulation of Visually Displayed Indexes. Int. Classification 7 (1) 21-24 (1980)
- 4) Craven, T. C. : NEPHIS Index Production from a General Concept Network. Proc. Am. Soc. Inf. Sci. Annu. Meet. 17, 180-182 (1980)
- 5) 郡司良夫 : PRECIS에 대하여 - 紹介의 試圖 大学図書館研究, 17 1-10 (1980)
- 6) Farradane, J. and Gulutzan, P : A Test of Relational Indexing Integrity by Conversion to a Permuted Alphabetical Index. Int. Classification 4(1), 20-25 (1977)
- 7) Neelameghan, A. and Gopinath, M. A. : Post-

- ulate-based Permuted Subject Indexing (POPSI) Libr. Sci. Slant Doc. 12 (3), 79-87 (1975)
- 8) 川村敬一 : Articulated Subject Index의 構造의 特性과 機械化 프로그램. 情報管理 24(5), 447-456 (1981)
- 9) Hutchins, W. J. : Languages of Indexing and Classification : A Linguistic Study of Structures and Functions. Stevenage, Herts, Peter Peregrinus Ltd., 1975, pp. 94-98.
- 10) Chomsky, N, 川本茂雄訳 : 言語와 精神, 東京, 河出書房新社, 1981, pp. 53-55
- 11) Svenonius, E. and Schmieder, H. F. : Current Issue in the Subject Control of Information. Libr. Quart. 47(3), 326-346 (1977)
- 12) Craven, T. C. : String Indexing : NEPHIS ; Introduction and Indexing. London, Ontario, School of Library and Information Science, Univ. of Western Ontario 1977{194p}
- 14) Farradane, J. : String Indexing : Relational Indexing; Introduction and Indexing. London, Ontario, School of Library and Information Science, Univ of Western Ontario, 1977 {240p}
- 15) Austin, D. and Verdier, V. : String Indexing PRECIS ; Introduction and Indexing. London, Ontario, School of Library and Information Science, Univ. of Western Ontario, 1977 {260}
- 16) Batty, D. : Book Reviews of 13) and 14) J. Doc. 35 (2), 159-161 (1979)
- 17) Batty, D. : Book Review of 15) J. Doc. 36(1), 93-94 (1980)
- 18) Armstrong, C. J. and Keen, E. M. : Microcomputer Printed Subject Indexes Teaching Package. Part 1 : Workbook for NEPHIS and KAC Part 2 : Manual for Teaching NPHIS and KWAC. Aberystwyth, Dyfed, College of Librarianship Wales, to be published in 1982
- 19) Craven, T. C. : A Government Publication Index Using Direct NEPHIS Coding of Titles. J. Inf. Sci. 2(1), 13-21(1980)
- 20) Fjerestad, L, and Craven, T. C. : NEPHIS Implementation Title Coding Manual. 2nd ed. London, Ontario, School of Library and Information Science. Univ. of Western Ontario. 1980 24p