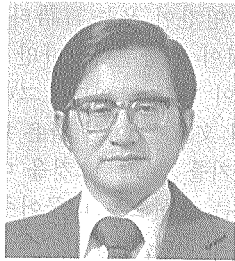


● 振興 컬럼

秒速翻譯의 時代가 오고있다.



金 貞 欽
高麗大 教授 / 理博

30年前에는 美國防省도 개발에 실패한 自動翻譯機

약 30年前인 1966年頃 美 펜타곤은 「러시아 語→英語의 自動翻譯機」 개발에 나선 일이 있었다. 아폴로 宇宙船으로 달世界까지 宇宙飛行士를 착륙시키고 무사히 歸還시키는 일이 한창일 때 이야기이다.

하버드大學의 研究팀을 중심으로 美國防省이 總力을 기울여 그 실현에 나섰지만 결국 이 프로젝트는 실패하고야 말았다. 그리고 이 失敗經驗이 원인이 되어 그후 한참동안 컴퓨터로 번역을 시킨다는 것은 무모한 짓이란 認識이 研究者들 사이에 퍼지게 되었다.

실패의 원인은 물론 컴퓨터技術이 當時에는 오늘날처럼 그리 높지 않았기 때문이다. 사실 지금의 技術로 보면 어린이 장난감처럼 보이는 당

시의 技術로써는 컴퓨터에게 「言語翻譯」을 시키겠다는 自體가 무리한 수작이었다.

그러나 그것으로 컴퓨터 科學者들이 降伏한 것은 아니었다. 그들은 계속 「言語翻譯」에 관한 꿈을 버리지 않고 컴퓨터 技術로지가 開花하는 것을 기다리면서 꾸준히 研究를 지속해왔다.

EC서 實用化된 翻譯機들

약 20年이 지나는 사이에 컴퓨터 技術은 눈부시게 발전하였다. 例컨대 컴퓨터 技術 발전의 測度의 하나인 IC의 集積度는 이 20年동안 거의 예외없이 每年 2倍씩 倍加되어 1979年末에는 1966年에 비해 약 1萬倍로나 늘어났다. 그리하여 처음에는 손톱크기의 基板 위에 트랜지스터 10개 정도를 모은 原始的인 集積回路(IC)에도 놀랐던 당시의 技術者들이 지금은 손톱크기의 마이크로 칩속에 컴퓨터의 中央演算處理部인 CPU를 넣을 수 있고 또 손톱의 1/3도 안되는 面積에 65536비트(64K 비트)의 記憶을 시키는 記憶裝置를 大量生産하는 데에도 별로 놀라지 않게 되었다.

그러자 그때까지 끈질기게 言語翻譯의 꿈을 키워온 컴퓨터 科學者들은 일제히 그 꿈을 실현시키려고 각 大學이나 研究所에서 言語翻譯機 개발의 경쟁에 나서기 시작했다.

특히 言語의 構造나 單語의 語幹이 같은 歐美 여러 나라에서는 英·獨·佛·伊 등등 隣接國語間의 翻譯機가 개발되어 實用化되기 시작했다. 특히 유럽 여러 나라로 構成된 유럽共同體인 EC에서는 言語翻譯機는 절실한 요청의 하나이기도 하였다. 즉 EC에서는 英·獨·佛·西 등의 7個國語로 된 公用語 사이에 迅速한 번역이 필요했고, 발표되는 모든 公文은 이들 7個國語로 번역되어 公示되어야만 하였다. 그 결과 1年間に 소비되는 번역費만도 16億弗이나 되어 EC全體 豫算의 數10%나 차지하여 무시못할 經費가 되어 있었다.

그러나 이제 EC(유럽共同體)의 7個 國語間의 42가지 [7(7-1)=42가지]에 達하는 번역기중 主宗을 이루는 英·獨·佛·伊 間의 번역기는 이미 實用段階에 들어가 經費節減에 지

대한 몫을 하고 있다고 한다.

또 英語使用地域과 佛語使用地域間에 항상 문제가 생기고 있는 캐나다에서는 英·佛語 自動翻譯機를 써서 氣象豫報를 表示하는 TV 畫面에서는 英語와 佛語가 동시에 表示되도록 되어 있다.

第3世代를 向한 自動翻譯 시스템

自動翻譯 시스템에는 3 가지 수준이 있다. 그것을 世代라 부른다면 그 첫째는 單語의 變換만을 하는 翻譯機로서 文法이 거의 비슷한 言語사이에서만 使用價値가 있으며 또 이런 경우에 있어서도 특히 간단한 文章이거나 技術關係 論文을 번역할 때 사용된다. 例컨대 재작년 韓·日 科學者들이 개발한 日→韓 翻譯機는 이 종류에 속한다. 또 文法은 갖지 않아도 單語만을 번역해 주는 기계, 例컨대 三星電子가 開發해서 인기를 얻은 英韓 및 韓英 포켓 電子辭典도 따지고 보면 이 카테고리에 들어간다고 보아야 하겠다.

둘째, 第2世代의 번역기는 文法的인 구조까지도 분석이 가능한 것으로서, 主語·動詞·目的語·補語 등등 文章의 기본적인 뼈대뿐만 아니라, 어느 單語가 어느 單語에 걸리는가 하는 修飾關係까지도 分析할 수 있는 능력을 갖고 있다.

이 方式으로는 文法만 맞으면 여러가지 엉뚱한 正譯아닌 誤譯도 나타날 가능마져 있다.

이 方式은 語幹이 거의 같고 文法이 다르다고는 해도 큼직하게는 서로 닮고 있는 言語群, 例컨대 Aryan 系에 屬하는 유럽 여러 나라의 言語群 사이의 翻譯에 쓰면 그 실력을 발휘할 수가 있다. 특히 기독교 文明, 그리스 文明, 로마 文明 등등의 영향을 받고 있는 유럽 여러 나라처럼 共通된 背景文化를 갖는 나라들의 言語사이의 번역에는 꼭 알맞는 번역기이다. 이런 경우 正譯率 95~98% 이상의 번역기를 만들기는 쉽다고 한다.

세째, 第3世代 翻譯機는 文法이나 語幹뿐만 아니라 背景文化 마저 전혀 다른 나라 말사이의

번역, 또는 文法이나 語幹이 같은 두 나라말 사이라도 文學的 香氣가 높은 詩나 小說 또는 廣告文 등을 번역하는 데 쓰이는 번역기이다.

이 水準의 翻譯機에서는 文法뿐만 아니라 文章의 意味까지도 완전히 이해해내는 능력이 있어야만 한다. 사실 말이란 것은 그 自體가 目的이 아니고, 그 말을 쓰는 사람의 머리 속에 들어 있는 어떤 概念을 말이라는 道具를 써서 表現한 것 뿐이다.

例컨대 「必死的」이라 해서 반드시 죽는다는 뜻은 아니다. 따라서 죽음에 直面하고 있는 환자를 「必死患者」라고는 하지않는다. 단지「힘껏 무엇인가를 한다는 뜻이다.

이와 같은 概念은 그 言語가 어떤 言語이던간에 인간이 쓰는 言語라면 그 概念體系는 共通된 것이 예상된다. 다만 각 나라의 言語마다 그 표현이 다를 뿐이다. 그래서 어느 한 나라 言語로 된 文章을 이와 같은 共通概念으로 일단 바꿔치기를 하고, 그것을 다른 나라 言語에서 사용되는 표현으로 바꾸어주면 완전한 意譯이 가능하게 된다.

그러나 이런 第3水準의 翻譯機는 현재의 컴퓨터技術으로써는 당장 만들어 내기가 힘들다. 다만 현재 世界의 여러 나라에서 開發中에 있는 第5世代(5G, Fifth Generatin)의 컴퓨터가 완성되면 비로소 가능해질듯 하다. 그리고 그 시기는 아마도 1990年代 後半에서 2000年代 初盤이 될듯 하다.

文法的으로는 완전한 誤譯

그러므로 현재로서는 第2의 水準인 第2世代 翻譯機가 追求되고 있고, 또 이미 商品化 되기 시작하고 있다.

그러나 이 第2世代의 翻譯機는 간혹 奇拔한 誤譯도 해낸다. 例컨대

「Time flies like an arrow.」

라는 英語의 번역이 그것이다. 正譯은 물론「歲月은 화살과 같다」이다. 그러나 文法的으로는 正譯이면서도 뜻으로는 어처구니 없는 다음과 같은 誤譯도 있을 수 있다.

「時間파리(複數)는 화살을 좋아한다.」

文法的으로는 하나도 잘못이 없으면서도 어처구니 없는 이와 같은 誤譯은 왜 나왔을까? 그것은 fly, flies란 單語에는 「난다」라는 動詞와 「파리」라는 名詞의 두 뜻이 있고, 마찬가지로 「like」란 單語에는 「—와 같은」이라는 前置詞와 「좋아한다」라는 動詞의 두가지 뜻이 있기 때문이다.

그러나 이런 경우 「시간파리」라는 昆蟲이 없다는 것은 明白하고, 「파리가 화살을 좋아한다」라는 말도 이상은 하다. 그러나 이런 종류의 「文法的으로는 完璧하면서도 뜻으로는 잘못된」 誤譯을 막을 길은 없다. 더구나 「시간파리」란 것이 中學校 英語先生님의 別名(時間이 되자마자 단 1秒도 틀리지않게 正確하게 教室에 들어오고, 또 강의를 할 때는 學生들을 한사람 한사람 指名해서 英語單語의 뜻이나 發言을 꼬치꼬치 캐물어, 파리처럼 돌아다니면서 學生들을 못살게 군다는 뜻에서 學生들이 붙인 닉네임일 수는 있다)이고, 그 先生님이 洋弓쏘기(archery)를 취미로 하고 있다면 이 誤譯은 誤譯이 아니고 名譯이 될 수도 있지 않겠는가. 「Time flies like an arrow.」가 무슨 뜻이나 先生님이 물었을 때 指定받은 學生이 일부러 모른체하고 「英語先生님은 洋弓을 좋아한다」라고 대답했다면 떠나갈듯 爆笑가 터진 教室 분위기와, 그리고 화가 잔뜩난 先生님의 怒氣어린 얼굴을 상상할 수 있지 않겠는가?

知識構造를 兼備한 第2世代의 번역기

이와 같은 誤譯아닌 誤譯을 막기 위해서는 어떻게 해야 하나? 그 방법은 컴퓨터 翻譯機에 「知識構造」를 넣어주는 일이다. 知識構造란 譯出된 文章이 현실적인 뜻을 갖는가 어떤가를 單語와 單語사이, 또는 그 나라 말의 습관 즉, 그 나라말의 慣用句 등을 일일이 조사하고, 檢索하고 대조하는 것을 뜻한다. 즉 컴퓨터에다 일종의 常識을 부여하자는 것이다.

예컨대 「먹는다」라는 動詞의 主語로서는 生物, 그중에서도 특히 動物만이 가능하다는 것을 조

사시키자는 것이다. 그렇기 위해서는 「人間·강아지·소·붕어·뱀·닭·새……」 등등은 「生物이다」라는 知識을 그 뜻과 함께 미리 주어져야 한다. 또 이 「人間」에는 「아버지·누나·소설가·컴퓨터기술자·운전수……」 등등이 들어 있다는 것도 가르쳐 주어야 한다.

마찬가지로 「먹는다」라는 動詞의 對象이 되는 目的語에 대해서도 「바다·山·江·大地……」 등등은 그 對象에서 除外해야만 한다. 이와 같은 意味構造, 概念構造 등의 分析이나 解析에는 무척 많은 量의 情報蓄積이 필요하기도하다.

그래서 이와 같은 知識構造를 具備한 컴퓨터 翻譯機같은 例컨대

「I bought a car with 6,000 dollars」

라는 英文을 「나는 6,000弗의 돈이 附錄으로 달린 自動車를 샀다」라고 번역을 하지는 않고 「나는 6,000弗로 車를 샀다」라는 正譯만을 내놓을 것이다.

사실 위의 경우 「6,000弗」이란 말은 「샀다」라는 動詞를 修飾하는 경우와, 車라는 名詞를 修飾하는 두가지가 文法的으로는 가능하다. 그러므로 「6,000弗로 車를 샀다」는 번역도 「6,000弗이 달린 車를 샀다」는 번역도 文法的으로는 옳바르다. 自動翻譯機로서는 文法에 의존하는 限 이 두가지를 구별할 수는 없다. 이 두가지 중 어느 것이 맞느냐 하는 것을 判斷해주는 것은 「6,000弗이라는 附錄이 달린 自動車」를 파는 販賣法은 없다는 것, 그리고 小型自動車의 값이 대략 5,000弗~8,000弗정도라는 사실 등등의 常識이다. 이 常識을 컴퓨터에게 提供해준다는 것은 말하기는 쉬워도 극히 힘든 일들이다. 왜냐하면 이런 常識은 數千萬, 數百億가지도 더 있어 그것만으로도 보통의 컴퓨터의 記憶容量을 훨씬 넘기 때문이다.

사실 科學者들이 꿈꾸고 있는 第5世代 컴퓨터를 쓴 第3世代 翻譯機라 할지라도 인간처럼 모든 것을 올바르게 번역해 줄 수는 없다. 어느 나라의 어느 文章이건, 그 文章의 뒤에는 社會의 常識이란 것이 들어 있고, 또 心理나 感覺에 관계되는 文學作品 등 컴퓨터가 번역하기는 무리한 것들이 많다.

그렇기는 하나, 이런 制約에도 불구하고 컴퓨터 사이언티스트들은 이런 理想的인 自動翻譯 시스템을 개발해 내기 위해 한 발자욱 한 발자욱 꾸준한 前進을 하고 있다.

벤처 비즈니스 會社가 開發해 낸 번역기

그런 과정의 産物의 하나로서 작년 6月 日本서는 「브라비스(Bravity)」라는 이름의 조그마한 벤처 비즈니스 會社가 「BRAVICE PAK 11/73/80(日本 돈으로 1,950萬円)」 및 「BRAVICE PAK 11/73/40(日本 돈으로 1,580萬円)」이라는 두 종류의 「日→英 自動翻譯機」를 내놓아 話題거리가 되어 있다. 이 自動번역기는 漢字와 日本가나(假名)가 섞인 「日本 國漢文」을 入力시켜 주면 1時間에 각각 3,000單語 및 2,000單語의 英文으로 번역을 해줄 수 있다고 한다. 이 번역기에 쓰이는 컴퓨터는 16비트의 미니컴퓨터이고, 그 크기는 높이 150cm, 幅50cm로서 記憶裝置에는 약 4萬單語의 英語 單語가 수록되어 있고, 별도로 사용자가 수시로 人力시켜서 등록시켜 쓸 수 있는 特用單語 記憶裝置가 마련되어 있다.

이 自動翻譯機의 正譯率은 原文의 종류에 따라 다르지만, 新聞이나 雜誌 등의 一般記事, 商品의 使用說明書, 技術資料, 技術論文이나 報告書 등 따위의 경우 약 85~92%라 한다.

이 정도의 正譯率이라면 충분히 글 전체의 뜻이 통한다고 한다. 그러나 만약 이 번역된 글을 英語專門家가 한번 읽고 약간의 손질만 한다면 손쉽게 100% 맞는 英譯이 가능하다고 한다.

물론 컴퓨터 自動翻譯機는 우수한 翻譯家보다도 그 翻譯의 質은 떨어진다. 그렇다고 해서 이 번역기가 불필요하다는 것은 아니다. 왜냐하면 우수한 翻譯家가 이 機械를 쓴다면 그는 약 400%~500%의 능률을 올릴 수가 있다고 한다. 따라서 이런 번역기 한대가 있다면 그는 번역기가 없었던 때보다 4배 내지 5배나 더 번역일

을 할 수가 있다는 것이 된다.

바람직한 韓→英 自動번역기의 開發

日本말은 韓國말과 그 文法이 사실상 같고, 또 韓字文化圈에 있는 만큼 單語 중 漢字起源의 것은 대부분이 그 뜻이 같다.

이런 點을 생각한다면 이 「브라비스 自動翻譯 시스템」은 약간의 손질만 한다면 손쉽게 「韓→英 自動번역기」로 改造가 가능하다. 따라서 우리도 하루 빨리 이 번역기를 導入해서 그 Knowhow를 분석해 내서 國產 「韓→英 自動翻譯機」를 개발해 내야만 한다. 그리고 이것이 끝나고 나면 다음에는 「英→韓 自動翻譯 시스템」을 개발해 내고 그 다음으로는 점차 韓國語와 프랑스어, 독일語, 스페인語 등등 世界 여러 主要言語의 번역기를 開發해 가면 된다. 듣건대 이 번역기는 商品化되기가 무섭게 美國防省이 이미 5대인가 6대를 구입해 갔다고 한다.

사실 情報化時代가 다가오는데 따라 國際의 情報交換의 量은 날로 늘어가고 있다. 그리하여 예컨대 우리나라 産業이나 文化를 소개하는 冊字나 우리나라의 外交政策 또는 觀光이나 올림픽 案内를 위한 各種 PR이나 情報를 재빨리 英譯이나 佛譯 또는 獨譯을 해서 外國에 提供할 필요가 있다. 그러려면 分秒를 다투어 大量의 情報文書를 번역해서 外國으로 보낼 필요가 있다. 이때 우수한 自動翻譯機 시스템이 있다면 막대한 量의 유리한 情報를 秒速으로 翻譯해서 外國 여러 나라에 흘려줄 수가 있다. 그렇게 하므로써 우리나라를 올바르게 外國에 소개한다는 것은 매우 중요할 뿐만 아니라, TV덤핑 문제, 輸出規制 문제, 韓國이 관련된 國內外 여러 事件이 날 때마다 우리의 正當한 主場表明을 재빨리 世界에 알려줌으로써 PR面에서 高地 獲保를 할 수가 있다.

그런 意味에서 自動翻譯機 시스템을 서둘러 開發해 주기를 진심으로 바란다.