

Preprint No. 62

COLLECTION OF ABSTRACTS OF PAPERS

INTERNATIONAL CONFERENCE

ON

******COMPUTATIONAL *******LINGUISTICS



C O L I N G

1969



RESEARCH GROUP FOR QUANTITATIVE LINGUISTICS

Address: *Fack Stockholm 40, SWEDEN*

ABSTRACT

"Machine Transcoding"

There are grounds for questioning whether translation done by computer should resemble closely, either in process or in product, translation done manually. Several workers in machine translation have proposed radical departures from the naively adopted goal of simulating manual translation mechanically. Because they saw a similarity between their mechanically feasible output and the pidgin languages that have arisen naturally in some parts of the world, they dubbed their proposals 'pidgin translation'.

This paper explores a more rigorous approach, which may be described as follows: texts in one language are encoded into words and other morphs taken from a second language, but this is done without disturbing the syntax of the source language. The term 'transcoding' is coined to denote that process. Transcoding results in a product in many ways similar to earlier types of pidgin translation (and incidentally more like natural pidgin languages than the latter were), but is founded on the principle that no information-bearing elements at all may be lost in the process. This principle is embodied in two strict guidelines:

1. Each and every grammatical morph in the original must be reproduced in the transcoded text by a code morph; the linear order of elements is not to be changed.
2. The transcoding dictionary (the codebook) must be so constructed as to preserve all the lexical contrasts inherent in the source vocabulary.

These guidelines are investigated and supplemented, and their effects are predicted with regard to (1) the output, (2) the user's ability to interpret transcoded material and to learn to interpret it better, (3) the possibility of employing transcoding to make rough drafts for the human professionals in translation agencies to revise.

Sample passages of transcoded Arabic and transcoded Chinese are presented, following which some comparisons are made between transcoding and pidgin translation, machine translation as practised or envisioned, and manual translation.

T.R. Hofmann,
Collège Militaire Royal,
St-Jean, Que.

Brian Harris,
Machine Translation Project,
Université de Montréal.

RESUME

Le transcodage automatique

On peut douter qu'une traduction automatique doive ressembler beaucoup à une traduction faite à la main, dans sa forme ou dans le processus qui le produit. Plusieurs d'entre ceux qui se sont attaqués aux problèmes de la traduction automatique ont proposé de s'écarter de façon radicale de ce but naïf qu'est la simulation automatique du processus de traduction humaine. Ayant vu quelques similarités entre leur sortie machine, qui de toute évidence représentait quelque chose de faisable sur ordinateur, et les langues naturelles dites créoles qui ont surgi en diverses régions du monde, ils surnommaient leur genre de traduction "traduction créolisée".

Cette communication fait le point d'une approche plus rigoureuse. Il s'agit de textes dans une langue source que l'on fait coder de nouveau en des mots et des morphèmes tirés d'une langue seconde, mais tout en conservant la syntaxe et même la sémantique de la langue source. D'après la présente communication, ce procédé s'appelle 'transcodage'. Le transcodage donc nous fournit un produit semblable à bien des égards à la traduction créolisée ci-dessus; d'ailleurs il se rapproche encore plus des créoles naturels que celle-ci. Cependant, il est basé sur un principe qui interdit formellement la perte, en cours du transcodage, de tout élément porteur d'information grammaticale ou sémantique. Ce principe se réalise par deux consignes:

1. Tous les éléments grammaticaux du texte source seront représentés dans la version transcodée; l'ordre linéaire des éléments ne doit pas être modifié.
2. Lors de la fabrication d'un dictionnaire destiné au transcodage, qui serait en effet une espèce de dictionnaire chiffré, on conservera soigneusement tous les contrastes ressortant de la structure du lexique de la langue source.

Nous explorons la portée de ces consignes, nous les complétons, et prévoyons leurs effets sur (1) la sortie (2) la possibilité pour l'utilisateur d'interpréter le matériel transcodé et d'apprendre à l'interpréter mieux (3) la possibilité d'employer le transcodage pour préparer des premières versions grossières que les spécialistes humains des agences de traduction pourraient améliorer.

Nous présentons des exemples d'arabe transcodé et de chinois transcodé, suivis de quelques comparaisons entre le transcodage, la traduction créolisée, la traduction automatique telle qu'on la pratique et telle qu'on l'envisage pour l'avenir, et la traduction manuelle.