



HAL
open science

Utilisation d'un système automatique de transcription du code sténographique en français, écrit pour le sous-titrage des émissions de télévision

F. Néel, C. Fluhr, C. Morel

► **To cite this version:**

F. Néel, C. Fluhr, C. Morel. Utilisation d'un système automatique de transcription du code sténographique en français, écrit pour le sous-titrage des émissions de télévision. *Revue de Physique Appliquée*, 1984, 19 (2), pp.133-138. 10.1051/rphysap:01984001902013300 . jpa-00245165

HAL Id: jpa-00245165

<https://hal.science/jpa-00245165>

Submitted on 4 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Classification
 Physics Abstracts
 01.90 — 43.70

Utilisation d'un système automatique de transcription du code sténographique en français, écrit pour le sous-titrage des émissions de télévision

F. Néel ⁽¹⁾, C. Fluhr ⁽²⁾ et C. Morel ⁽³⁾

⁽¹⁾ Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, 91406 Orsay, France

⁽²⁾ Université Paris-Sud, 91405 Orsay, France

⁽³⁾ Centre Commun d'Etudes de Télédiffusion et Télécommunications, 35000 Rennes, France

(Reçu le 3 juin 1983, révisé le 12 septembre, accepté le 9 octobre 1983)

Résumé. — Le système de télé sous-titrage numérique ANTIOPE, permet aux malentendants d'avoir une meilleure compréhension des émissions de télévision, sans pour autant gêner l'ensemble des téléspectateurs. Ce nouveau service, dont l'usage devrait se développer rapidement, pose des problèmes de production, difficiles à résoudre par les moyens classiques :

1. Le sous-titrage des émissions produites en direct.
2. Le coût élevé de production des sous-titres.

Une solution séduisante serait d'utiliser un dispositif de reconnaissance de la parole comme organe de saisie ; malheureusement l'état de l'art actuel ne le permet pas.

Une autre méthode semble possible dès à présent ; elle utilise la saisie par sténographie, et nécessite l'étude du système de transcription par ordinateur pour passer du code sténographique (de type phonétique) vers le français écrit.

Cette méthode fait l'objet d'études menées conjointement par le CCETT, le LIMSI, la SYSTEX, et la Sténographie GRANDJEAN. Le principe de cette méthode et les problèmes posés sont exposés dans cet article. D'autres applications qui peuvent en découler sont également citées.

Abstract. — The ANTIOPE Teletext system used for TV subtitling, enables hearing-impaired people to have a better understanding of TV, without troubling the majority of the TV watchers. This new service, the use of which should spread out shortly, gives production problems that are difficult to solve through classical means :

1. The subtitling of live TV programs.
2. The high cost of subtitled programs.

An attractive answer would be to use a speech recognition system. Unfortunately, it is not technically possible, at the present state of the art.

Another method which seems already possible, is the use of stenotype entry. It requires the study of computer aided transcription (CAT system), in order to translate stenotype code to written French. This method formed the subject of cooperative research lead by CCETT, LIMSI/CNRS, SYSTEX and « Stenotypie GRANDJEAN ».

The main principle of this method and the underlying problems are described in this article. Others applications are also quoted.

1. Introduction.

Il est indispensable, pour répondre aux besoins des trois millions de malentendants français, d'utiliser le sous-titrage des émissions de télévision. Grâce au système de télé sous-titrage numérique utilisant le système de Télétex ANTIPOPE DIDON [1, 2], il est possible d'adresser les sous-titres à un public déterminé, sans gêner l'ensemble des téléspectateurs. Ce nouveau service, qu'on pourrait qualifier de sous-

titrage à accès sélectif (« closed captioning » en anglais), nécessite chez l'utilisateur l'usage d'un décodeur associé au récepteur de TV (décodeur ANTIPOPE). Il est appelé à se développer en France, où les premières expérimentations sont en cours.

La figure 1a montre quels types d'équipements sont utilisés actuellement pour créer et diffuser les sous-titres dans le cas où l'on diffuse des émissions enregistrées sur bande magnétique ou sur film.

La figure 1b montre les avantages qu'offre l'utili-

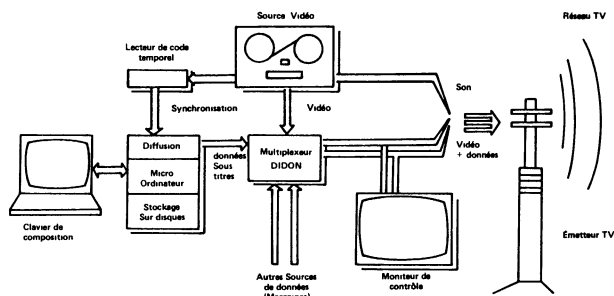


Fig. 1a. — Source de sous-titrage ANTIOPÉ conventionnelle.

[Commonly-used ANTIOPÉ subtitling source.]

2. La saisie sténotypique intégrée dans une source de sous-titres peut fournir une solution.

Il faut, pour préciser ces problèmes, donner une estimation du travail nécessaire à la création des sous-titres : avec une source conventionnelle (utilisant un clavier alpha-numérique), il faut de 20 à 30 h de travail pour une émission d'une heure. Le sous-titrage en direct n'est pas possible dans ces conditions, sauf si le contenu de l'émission est connu précisément par avance. On se contente donc de sous-titrer les émissions enregistrées. Les objectifs de T.D.F. (Télédiffusion de France) et des trois chaînes de télévision sont à court terme, et en utilisant des sources conventionnelles (Fig. 1a), de sous-titrer de 1 h à 3 h d'émission par semaine et par chaîne en fin 83, puis 5 h fin 84.

Pour obtenir un gain de temps à la saisie de la parole, on peut envisager l'une des solutions suivantes :

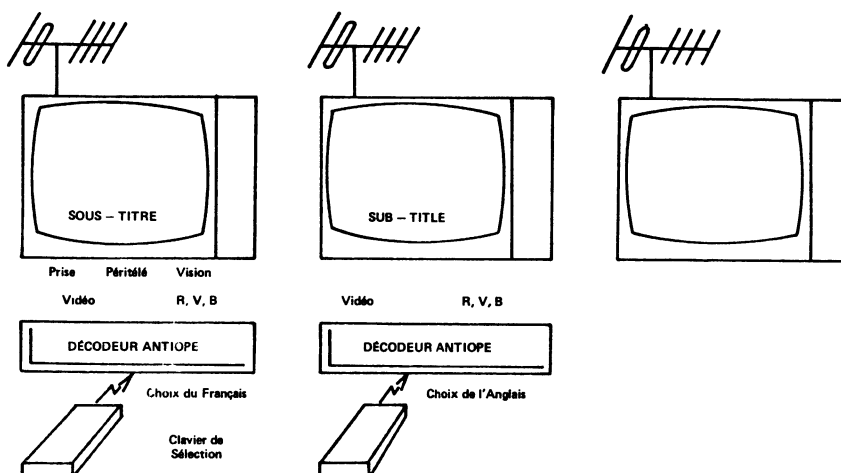


Fig. 1b. — Réception d'une émission sous-titrée par ANTIOPÉ en 2 langues.

[Reception of a 2-language subtitled broadcast, using ANTIOPÉ Teletext.]

sation d'un décodeur ANTIOPÉ par rapport au sous-titrage classique (où les sous-titres sont inscrits de manière visible et définitive dans l'image) :

— aucune gêne pour le public qui entend la langue parlée,

— possibilité d'adresser différentes versions sous-titrées pour répondre aux besoins d'auditoires spécifiques (sourds, malentendants, communautés linguistiques différentes),

— possibilité de recevoir également des magazines d'informations transmis par le même système DIDON-ANTIOPÉ : tels que ceux diffusés actuellement sur les chaînes nationales de télévision : météo, A 2, état des routes, etc...

Cependant on est très vite confronté à des problèmes de production dont les deux suivants sont mal résolus :

1. Sous-titrage des émissions produites en direct (par exemple les journaux d'information télévisés).

2. Minimisation des temps et des coûts de production des sous-titres.

— Utiliser un système de reconnaissance de la parole (Fig. 2) : la parole émise par le locuteur serait convertie immédiatement en texte écrit par ce système, de manière entièrement automatique.

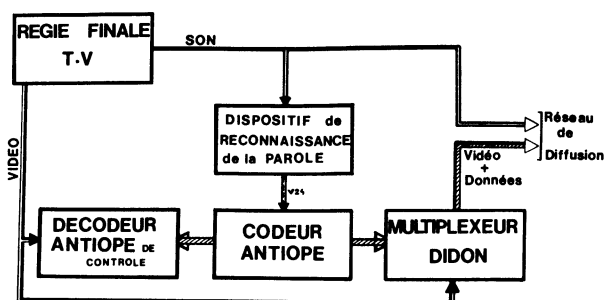


Fig. 2. — Sous-titrage par la reconnaissance de la parole.

[Subtitling source using speech recognition.]

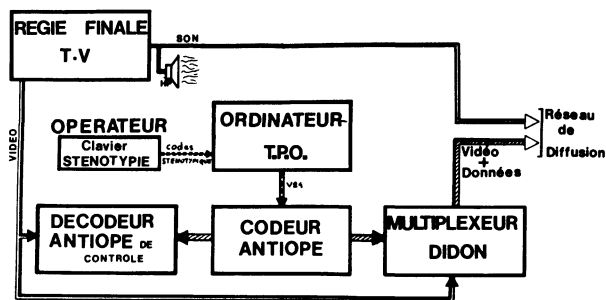


Fig. 3. — Sous-titrage par la saisie sténographique et la Transcription Par Ordinateur (TPO).

[Subtitling source using stenotype input and Computer Aided Transcription (CAT).]

— Utiliser un clavier à entrée phonétique (Fig. 3) : un opérateur spécialisé effectuerait en direct la saisie sténographique, et le code sténographique serait converti automatiquement en texte écrit.

Dans les deux cas, on est conduit à étudier la transcription de phonèmes (symboles représentant les sons — voyelles, consonnes — de la langue) en graphèmes (lettres du français écrit). Si la première solution semble idéale puisque n'utilisant aucun opérateur humain, elle n'est pas encore réalisable dans le contexte d'un vocabulaire étendu, d'un flot continu de parole et de plusieurs locuteurs. De nombreuses études sont consacrées à ce problème de par le monde, mais il serait utopique d'en espérer prochainement une solution générale. En outre, il n'est pas certain que la transcription pure et simple de la langue parlée en texte écrit fournirait des sous-titres parfaitement intelligibles.

La deuxième méthode paraît plus rapidement applicable, car elle fournit des codes phonétiques pratiquement déterministes en utilisant la méthode de sténographie GRANDJEAN [3]. Cette méthode, utilisée couramment en saisie des discours de conférence, permet de suivre en temps réel les locuteurs les plus rapides (280 mots/min). Des systèmes de transcription automatique par ordinateur (TPO en français ou CAT en anglais) existent aux U.S.A. pour la méthode sténographique anglophone « PALANTYPE » [4, 5]. L'un de ces systèmes utilisé près de Washington par NCI (National Captioning Institute) réalise depuis près d'un an, la transcription en temps réel des Actualités (ABC-TV's « World News Tonight »). Les performances sont encourageantes.

Il n'existe pas encore de produit équivalent pour la méthode francophone GRANDJEAN. Le but de notre étude est donc la réalisation d'un système de TPO (Transcription par Ordinateur) pouvant fonctionner en direct, maillon compris entre le clavier sténographique et un codeur ANTOIPE (Fig. 3). Une étude analogue est menée par IBM-France avec une approche différente du problème.

Les principales caractéristiques de la méthode de sténographie GRANDJEAN permettent d'appréhender les difficultés à surmonter dans l'étude de l'automate de TPO.

— Le clavier est composé de 21 touches représentant des sons élémentaires : consonnes initiales, voyelles, consonnes finales (Fig. 4). Le nombre de codes sténographiques est inférieur au nombre de phonèmes représentant la langue française, ce qui entraîne de nombreuses ambiguïtés de nature phonétique.

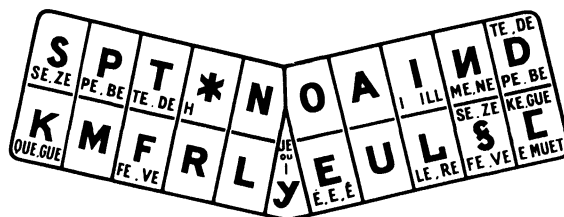


Fig. 4. — Clavier de la Sténographie GRANDJEAN.

[GRANDJEAN Stenotype keyboard.]

— L'écriture se fait syllabe par syllabe (on peut enfoncer simultanément plusieurs touches); il n'y a pas de séparation entre les mots (Fig. 5). Il en résulte des ambiguïtés lexicales et syntaxiques.

<p>LE I M P R E I K M E I F O U A N O U A F E A T R E A S T R E E S S L A U L S T N O I I P N O U I I T N O U I I T R E E I I S N O A M M T R O A M S F L A A I S T L E I I S T O E I I T F* E U D P R L A A I P T T E I I K P R L A I S P E U</p>	<p>Les imprimés que vous nous avez adressés sur la sténographie nous ont intéressés en nous montrant la facilité de son étude (virgule) la rapidité de la prise qui peut</p>	<p>P T E A S T E U I I S M O A I I M L A I I D P L O A S P L O I I I T T E I I P T R A I I T O U A I I D * L A I I D C * * N O U A I I F N O A I I T N O E I I S K T O I I K T L O U I I M R L O I I D S K T E I I S T T E E I I P T N O I I P E</p>	<p>dé- pa (s) ser deux cents mots à la minute et la possibilité de prendre en toutes langues Point à la ligne) Nous avons noté au (s) si que tout le monde relit ce qui a été sténographié</p>
--	--	--	--

Fig. 5. — Bande sténographique et sa transcription.

[Stenotype paper with its corresponding transcription.]

Toutes ces ambiguïtés, dont des exemples seront donnés plus loin, sont normalement levées par l'opérateur quand il relit son ruban sténographique et le convertit en texte orthographique. Il utilise pour cela sa connaissance de la langue française à tous les niveaux (phonétique, syllabique, lexical, syntaxique, sémantique), ainsi que sa connaissance de la situation et des conditions dans lesquelles la saisie a été faite.

3. La transcription sténographique-orthographique.

On peut s'attendre à ce que des processus linguistiques aussi complexes soient extrêmement difficiles à automatiser, et requièrent au moins un lexique complet du français, et une méthode d'analyse syntaxique permettant de traiter la langue naturelle en temps réel. Mais, en premier lieu, il faut savoir d'où viennent les principales ambiguïtés, et savoir si elles peuvent être levées en utilisant les outils dont on dispose actuellement.

Dans la mesure où nous ne pouvons nous appuyer sur les résultats obtenus en anglais, d'une part, parce que les structures linguistiques en sont différentes et, d'autre part, parce que le système de codage PALANTYPE fournit une information moins ambiguë aux niveaux phonétiques et syllabiques, il nous a semblé nécessaire d'entreprendre une étude de faisabilité avant de passer à la réalisation d'une maquette.

3.1 ANALYSE DES RÉSULTATS DE L'ÉTUDE DE FAISABILITÉ.

3.1.1 *Les conditions de l'expérimentation.* — Un échantillon de 22 750 mots, représentant 8,75 % du lexique entier, a été extrait d'un dictionnaire de plus de 200 000 formes complètes utilisé par le système SPIRIT [6] qui permet l'interrogation de banques de données textuelles en langage naturel (base documentaire juridique...). Par « formes complètes », nous entendons que les mots et adjectifs, formes conjuguées à tous les temps et modes pour les verbes, se trouvent dans le dictionnaire.

Ces mots ont été traduits automatiquement en code sténographique, en passant par l'intermédiaire d'un module de transcription graphèmes-phonèmes [7] : « pagination » est traduit /pazinasjõ/ en phonétique puis PA-YI-NA-SYON en sténotypie. De même « franchit » donne /frãʃi/ puis FRAH-SKI etc.

Un programme de statistiques a ensuite déterminé le nombre des ambiguïtés de traduction (plusieurs orthographes pour un même mot sténographique) : par exemple, les sons /e/ et /ɛ/ sont traduits par un seul caractère sténographique E qui correspond à plus d'une dizaine de mots de la langue française : ai, aie, aies, ait, aient, est, es, et, haie, haies, hais, hait... Les ambi-

guïtés qui ne sont pas issues de problèmes d'accord (genre, nombre, temps, mode, personne) ont seules été répertoriées : dans l'exemple précédent, seules cinq formes ont subsisté : ai (verbe avoir), es (verbe être), et (conjonction), haie (substantif), hais (verbe haïr).

3.1.2 *Les résultats de l'expérimentation.* — Pour les 22 750 formes orthographiques traitées, le système a relevé 10 322 formes sténographiques soit un nombre moyen d'ambiguïtés de 2,2 formes orthographiques pour une forme sténographique.

Les ambiguïtés relevées peuvent être réparties en plusieurs classes :

1. *Ambiguïtés dues aux caractéristiques phonétiques de la langue française* (96,16 % des ambiguïtés)

A) Homophones hétérographes : une prononciation se réfère à plusieurs formes orthographiques possibles :

a) accord genre/nombre (substantifs, adjectifs, conjugaison verbale) (94,67 % des ambiguïtés)

chateau/châteaux
vieil/vieille/vieilles
mange/manges/mangent etc.

b) /s/ (son s) qui peut s'écrire -ss-, -c-, -t- (0,95 % des ambiguïtés)

épurations/épuration
aimassions /émacions

c) /e/ (son é) qui peut s'écrire -ai-, -é-
/ɛ/ (son è) qui peut s'écrire -ai-, -è- ou -e- suivi de 2 consonnes (0,08 % des ambiguïtés)

adulaire/adulèrent
épièrent/épierre

d) /ã/ (son an) qui peut s'écrire -an-, -en-, -aon-, -aen-, ... (0,3 % des ambiguïtés)

adhérant/adhérent

B) Défauts introduits par le programme de traduction phonétique :

a) qui ne donne qu'une forme phonétique possible pour une forme orthographique et ne traite donc pas les homographes hétérophones : (0,16 % des ambiguïtés) :

— mots terminés en -tions- ; t se prononce /t/ ou /s/

options se prononce /ɔpsjõ/
/ɔptjõ/

— mots terminés en -ent- (se prononce /ə/ ou /ã/)

négligent se prononce /negliz/ ou
/neglizã/

b) qui ne traite pas les noms propres ou certains noms communs d'origine étrangère, à la prononciation particulière.

2. *Ambiguïtés dues au code sténographique (2,8 % des ambiguïtés).*

— 1 seul caractère pour 2 ou 4 consonnes possibles en fin de syllabe (I, J, D, L, S, [])

accidentel/accidentèrent

même code :

A[-SI-T*AI-TEL

— 1 seul caractère pour -g- ou -gi- (y)

aménageons/aménagions

même code :

A-ME-NA-YOM

3. 1,04 % des ambiguïtés sont dues à des erreurs dans les programmes ou dans le dictionnaire.

3.2 DISCUSSION DES RÉSULTATS ET SOLUTIONS PROPOSÉES.

1. *Comment traiter le problème des ambiguïtés et du découpage de la chaîne en mots ?*

Il apparaît de manière évidente que les ambiguïtés donnant pour une même forme sténographique des mots différents de la langue sont en très petit nombre. A eux seuls, ils ne justifieraient pas la mise en place d'un traitement élaboré.

Par contre, la quasi-totalité des ambiguïtés porte sur les accords pour un même mot. La fréquence de telles ambiguïtés dans un texte devrait être de l'ordre de un mot sur deux. Ceci rend toute application de transcription impossible sans un traitement efficace de ce problème.

Le traitement de telles ambiguïtés ne peut se contenter d'une analyse syntaxique élémentaire (comme celle permettant de lever des ambiguïtés substantif-verbe, adjectif-verbe, etc.). Il faut reconnaître des relations de type sujet-verbe, sujet-attribut, en plus de relations plus évidentes telles que article-substantif, préposition article-substantif, adjectif-substantif.

Ce problème est propre au français, ce qui explique pourquoi les solutions assez élémentaires réalisées aux U.S.A. ne sont pas transposables chez nous.

Il serait possible, dans le cadre de la seule application « sous-titres » de réaliser un système sans analyse syntaxique, mais cela donnerait une qualité de sous-titres très mauvaise et obligerait de former des sténographes à un nouveau type de codage. Cette solution ne peut être retenue.

On peut remarquer que l'analyse syntaxique devient obligatoire dès que l'on se pose le problème du découpage de la chaîne sténographique en mots. En effet, les ambiguïtés de découpage sont levées à un premier niveau par le lexique, mais il subsiste des ambiguïtés qui doivent être résolues par l'analyse syntaxique. En effet, la phrase « le chaland serre la souris » renvoie à plusieurs découpages lexicaux possibles : le chat lent sert la souris, le châlè enserre lasse ou ris, le châlè l'en cerf là sourd rit, etc. qui ne sont pas tous syntaxiquement corrects.

Dans ce cas aussi, certaines ambiguïtés de découpage ne pourront être résolues seulement par l'analyse syntaxique. Il sera nécessaire de prendre en compte les fréquences lexicales en fonction du domaine de l'application.

On ne pourra évaluer l'importance de ces ambiguïtés qu'après expérimentation d'une transcription utilisant au moins un niveau d'analyse syntaxique.

2. *Problèmes posés par l'application sous-titrage en temps réel.*

En tenant compte du retard apporté par le sténographe, retard qu'on peut évaluer de l'ordre de 1 à 2 s, et en sachant qu'un décalage global entre les paroles et les sous-titres correspondants de plus de 10 s rendrait incompréhensibles les sous-titres d'une émission en direct, il apparaît que le retard imputable à la transcription par l'ordinateur, devra toujours être inférieur à 5 s. Cette contrainte importante montre que cette application exigera un ordinateur de forte puissance.

3. *Suite prévue pour cette étude.*

Une fois réalisée une maquette de TPO sur gros ordinateur, des tests en vraie grandeur permettront d'évaluer les performances du système en temps réel : taux résiduel de fautes, retard apporté. Un test d'évaluation subjective de l'acceptabilité d'un service de sous-titrage des émissions en direct sera mené auprès des malentendants.

En fonction des résultats, on envisagera la mise au point d'un appareillage spécialisé utilisant un micro-ordinateur de forte puissance.

4. Conclusion.

En s'appuyant sur les travaux déjà réalisés dans le domaine de la transcription phonétique-orthographique, *il semble nécessaire d'envisager un système de traduction sténographique-orthographique intégrant des traitements syntaxiques très profonds. Les ambiguïtés qui subsisteront seront de nature sémantique et en très petit nombre.*

Une autre retombée immédiate de ces travaux concerne la transcription des textes de conférences, saisis en sténographie. Cette application ne nécessite pas un traitement en temps réel, et peut faire usage d'un microordinateur de moyenne puissance.

On aura aussi la possibilité, grâce à la saisie sténographique automatisée, de disposer de textes sur support informatique, et de constituer des bases de données textuelles facilement consultables (comptes-rendus de conseils, jurisprudence, etc.).

A plus long terme, cette recherche est une étape préalable à la réalisation d'une machine à écrire vocale (« dictée automatique ») [8]. Les méthodes d'analyse employées dans la transcription sténographique sont en effet très proches de celles qui sont à l'étude dans le domaine de la reconnaissance automatique de la parole.

Bibliographie

- [1] DIDON ANTIOPE : Spécifications techniques. La Radiodiffusion-Télévision. Document n° 1, 1982.
 - [2] SECHET, C., Sous-titrage des émissions de télévision par le système ANTIOPE. Congrès International sur les systèmes et services nouveaux de télécommunications. 24-26 novembre 1980 Lièges.
 - [3] Méthode de Sténotypie GRANDJEAN. Grandjean Editeur.
 - [4] JORPELAND, M. S., Court reporter helps make the silent screen speak. NSR. April 1982.
 - [5] KARLOVITS, J. R., Equal justice, The Deaf and court reporting. NSR. July 1982.
 - [6] ANDREWSKY, A., BINQUET, J. P., DEBILL, F., FLUHR, C., HLAL, Y., LIENARD, J. S., MARIANI, J., POUDEROUX, B., Les dictionnaires en formes complètes et leur utilisation dans la transformation lexicale et syntaxique correcte de chaînes phonétiques. 10^e, JEP, 1^{er} juin 1979, Grenoble.
 - [7] PROUTS, B., Traduction phonétique de textes écrits en Français. 10^e, JEP, 30 mai, 1^{er} juin 1979. Grenoble.
 - [8] ANDREWSKY, A., FLUHR, C., LIENARD, J. S., MARIANI, J., NEEL, F., Méthode d'apprentissage pour la reconnaissance automatique de textes lus à voix haute. Symposium Franco Soviétique sur la parole, GALF, Grenoble, octobre 1981.
-