

Le sous-titrage des émissions de télévision  
à l'usage des mal-entendants

TV captions for the deaf

Bernard Merialdo, Anne-Marie Derouault

Centre Scientifique IBM-France, 36 Avenue Raymond Poincaré 75116 Paris  
FRANCE

---

**RESUME**

Le développement des techniques de communication et de traitement de l'information offre de nouvelles applications pour l'aide aux handicapés. Cet article s'intéresse aux possibilités de sous-titrage en temps réel à l'usage des sourds et mal-entendants. Ce sous-titrage est particulièrement intéressant pour les émissions de télévision. De tels systèmes sont déjà expérimentés en Angleterre et aux Etats-Unis. Grâce au système de transcription automatique de la sténotypie en Français réalisé par le Centre Scientifique IBM-France, des expériences équivalentes sont désormais possibles en France.

**SUMMARY**

The development of information processing and communication techniques opens new applications in the aid to the handicapped. This paper focuses on the feasibility of real-time captioning for deaf people. Captioning is specially interesting for TV broadcasts. Such systems have already been experimented in Great-Britain and in the USA. A stenotypy-to-French transcription system has been developed by the IBM-France Scientific Center, so that equivalent experiments are now possible for French.



## LE PROBLEME DU SOUS-TITRAGE

Les sourds et mal-entendants, de par leur handicap, ont de grandes difficultés pour bénéficier des moyens de communication basés sur la parole: dialogue, radio, télévision, téléphone. De nombreux systèmes existent pour remplacer le canal acoustique par un canal visuel, afin de leur donner la possibilité d'utiliser ces moyens, et ainsi leur permettre de mieux s'intégrer dans la société. Ces systèmes sont par exemple: la lecture sur les lèvres, le langage des signes, ...

La plupart de ces systèmes demandent un effort d'apprentissage important. Souvent, ils s'appuient sur un code qu'il faut connaître et maîtriser. Ces codes créent parfois des problèmes de communication entre le monde des sourds et celui des entendants.

Un moyen commode est de pouvoir utiliser comme support visuel la transcription écrite du discours, c'est-à-dire réaliser un sous-titrage textuel. Un tel sous-titrage peut s'appliquer aussi bien pour la télévision, le cinéma, que pour des conférences, des discussions ou des communications téléphoniques.

Le problème est d'arriver à le produire.

## LA SITUATION EN FRANCE

La réalisation des sous-titres pour les émissions de télévision pose deux problèmes:

- comment fabriquer le texte,
- comment l'intégrer à l'image.

Le système de télétexte ANTIOPE développé par le CCETT (Centre Commun d'Etudes de Télédiffusion et Télécommunications) résout de façon efficace le deuxième problème. Il permet d'émettre des sous-titres qui ne seront reçus que par les personnes possédant un décodeur ANTIOPE. Celui-ci superpose le texte à l'image normale de l'émission. Ce procédé permet de ne pas perturber la diffusion normale des émissions.

Mais la fabrication des sous-titres exige un long travail. Il faut d'abord avoir le scénario de l'émission, concevoir les résumés, les intégrer et les synchroniser avec l'image. Ce processus demande au mieux 30 heures, et souvent de 50 à 100 heures de préparation pour une heure d'émission. C'est donc un processus onéreux et peu commode. D'autre part, il limite l'application des sous-titres à des émissions pré-enregistrées, et donc exclut celles en direct.

## LE SOUS-TITRAGE PAR LA STENOTYPIE

Le sous-titrage des émissions en direct demande la production en temps réel de la transcription du discours. Dans l'état actuel des connaissances, il est difficile d'envisager de faire un résumé en temps réel, aussi la seule solution est de produire l'intégralité du discours.

La dactylographie ne permet pas une vitesse suffisante.

S'il existait un système de reconnaissance de la parole assez performant (multi-locuteurs, grand vocabulaire, faible taux d'erreurs, temps réel), le problème serait résolu. Malheureusement, un tel système n'existe pas encore.

Cependant des solutions commencent à apparaître utilisant la transcription automatique de la sténotypie.

## LA STENOTYPIE

La sténotypie est une méthode de codage pseudo-phonétique de la parole qui autorise une saisie en temps réel (en fait il y a plusieurs méthodes). Elle est basée sur l'utilisation d'un clavier (de 21 à 29 touches selon les pays) où l'on appuie simultanément sur plusieurs touches de façon à composer une syllabe pseudo-phonétique (sténogramme). Le discours est ainsi codé par une suite continue de sténogrammes.

Des systèmes de transcription automatique existent déjà en Angleterre, aux USA et maintenant en France. Ils produisent automatiquement le texte écrit au fur et à mesure que les sténogrammes sont frappés, et permettent donc la production en temps réel de sous-titres textuels.

## LES EXPERIENCES EN ANGLETERRE ET AUX ETATS-UNIS

En Angleterre, c'est la méthode Palantype qui existe. Un système de transcription a été développé par l'Université de Southampton [1]. Il contient un dictionnaire de 1200 mots. A partir des sténogrammes frappés, le système recherche si des mots du dictionnaire apparaissent. Si oui, leur orthographe est affichée. Pour les sténogrammes qui ne correspondent à aucun mot, le système compose une transcription orthographique qui représente l'écriture la plus probable de ce son. Le texte résultant est donc un mélange de mots et de syllabes plus ou moins correctement orthographiées (mais en général lisibles). Une expérience de sous-titrage a été effectuée à l'occasion des fêtes du mariage princier. Ce système est également utilisé par un député britannique sourd [2] qui, grâce à lui, suit et participe aux débats du parlement.

Aux Etats-Unis, c'est la méthode Stenograph qui est employée. L'utilisation systématique de la sténotypie dans les tribunaux a contribué au développement de nombreux produits de Transcription Assistée par Ordinateur ('CAT' pour 'Computer Aided Transcription'). Ces systèmes sont toujours basés sur de gros dictionnaires, la plupart du temps construit au fur et à mesure par l'utilisateur (pour tenir compte des variations de la méthode). De nombreuses expériences de sous-titrage ont été menées, dans des classes [8] [9] ou au cours d'émissions de télévision [3] [4].



## LE SYSTEME TASF DE TRANSCRIPTION DE LA STÉNOTYPIE

En France, c'est la méthode de sténotypie Grandjean qui est utilisée.

Le système expérimental TASF (Transcription Automatique de la Sténotypie en Français) a été réalisé par le Centre Scientifique IBM France [6] [5]. C'est le premier système de ce type. Il permet de transcrire une suite de sténogrammes en un texte Français. Il est basé sur un grand dictionnaire sténotypie-Français (140.000 formes fléchies), et sur un modèle probabiliste du Français permettant de choisir entre les orthographes possibles. Contrairement à l'Anglais, le Français nécessite un modèle de langage en plus du dictionnaire, à cause du plus grand nombre d'homonymes (par exemple singulier/pluriel ou conjugaisons) et de mots courts [7].

Bien que la destination première de ce système soit orientée vers une transcription en différé, ce qui correspond aux besoins des sténotypistes de conférence, les algorithmes utilisés permettent une utilisation en direct. En effet, le modèle probabiliste basé sur des contraintes locales fait que la transcription a un retard limité sur les sténogrammes saisis. Ce décalage varie entre 4 et 6 mots, selon le nombre d'ambiguïtés rencontrées.

Une version du système tourne sur un ordinateur personnel IBM PC-XT. La vitesse de transcription est alors de 90 mots/mn, ce qui est insuffisant pour le temps réel (entre 160 et 200 mots/mn). Par contre, une autre version tourne sur un IBM 3031 avec une vitesse de transcription de 300 mots/mn, ce qui dépasse largement les possibilités d'élocution de n'importe quel locuteur.

Le taux d'erreurs (pourcentage de mots incorrectement transcrits) se situe entre 8 et 9%. Il faut rajouter en plus les fautes de frappe de la sténotypiste, ainsi que les mots qui ne figureraient pas dans le dictionnaire (noms propres et termes techniques). Ces erreurs sont dues au codage pseudo-phonétique qui introduit des ambiguïtés au niveau syntaxique ('parle' vs 'par le'), sémantique ('chêne' vs 'chaîne'), ou pragmatique ('des jours néfastes' vs 'des journées fastes').

En ajoutant les délais de frappe de la sténotypiste, de transcription, et de décalage de la transcription par rapport à la frappe, on voit que le système est capable de fournir le texte avec seulement quelques secondes de retard par rapport à la parole, ce qui remplit la condition de temps réel.

## LISIBILITE DES SOUS-TITRES

La lisibilité est liée à deux facteurs:

1. le taux et le genre des erreurs du texte produit. Certaines erreurs sont peu gênantes pour la compréhension du texte. C'est le cas par exemple de certaines fautes d'accord qui sont facilement rectifiées à la lecture. D'autres comme les erreurs de frappe, rendent plus délicate la relecture du texte, car elles segmentent souvent un mot long en plusieurs mots courts, et demandent un effort mental de reconstruction du mot pour une compréhension correcte.

2. la procédure d'apparition du texte à l'écran. Cela met en jeu d'une part la vitesse de défilement. Une lecture rapide sera certainement plus facile pour une personne devenue sourde, qui a pu acquérir une bonne connaissance du langage, que pour une personne sourde de naissance, pour laquelle les connaissances linguistiques sont plus difficiles à acquérir.

D'autre part la disposition du texte sur l'écran et son évolution dans le temps joue un rôle important dans la lisibilité. Plusieurs dispositions sont possibles, dont certaines sont mieux adaptées à un défilement rapide du texte.

La réponse à ces deux questions ne peut se faire qu'au travers d'expériences avec des personnes sourdes.

## LES METHODES DE DEFILEMENT

Parmi les méthodes possibles de défilement de texte, nous en retiendrons 3 qui nous semblent les plus caractéristiques:

1. par bloc

C'est la méthode employée actuellement dans les sous-titres (aussi bien à l'usage des mal-entendants que pour le cinéma). Elle consiste à découper le texte en blocs successifs. Chaque bloc reste un temps minimum à l'écran, puis laisse la place au bloc suivant.

2. par décalage

C'est une méthode utilisée souvent lors des génériques. Elle consiste à décaler le texte en effaçant les informations les plus anciennes de façon à laisser de la place pour les nouvelles informations. Les nouvelles informations apparaissent toujours à la même place sur l'écran. Le décalage peut être vertical ou horizontal, continu ou discret, au niveau du caractère, du mot ou de la ligne.

3. par recouvrement

Cette méthode consiste à placer les nouvelles informations par dessus les informations les plus anciennes de l'écran (sans décalage du texte). Elle peut s'appliquer au niveau du caractère, du mot ou de la ligne. Par exemple, au niveau de la ligne, l'ordre d'apparition sur un écran de n lignes numérotées 1,2...n sera 1,2,...n,1,2...n,1... La n+1ème ligne du texte viendra s'afficher par dessus la première. L'écran contiendra alors les lignes n+1,2,...n.

Nous pensons que les méthodes 2 et 3 sont les plus adaptées à une vitesse rapide de défilement.

En effet, considérons le plan *mot du texte/temps*. Dans ce plan on peut marquer les points correspondants à la partie du texte visible à chaque instant, ce qui forme le *champ de vision*. Pour la méthode par bloc, on obtient un graphe montré dans la figure 1, pour les méthodes par décalage ou recouvrement, un graphe montré dans la figure 2 (par simplicité, on n'a pas discrétisé le graphe).



La pente de la droite dans chacun des graphes représente la vitesse moyenne de défilement du texte.

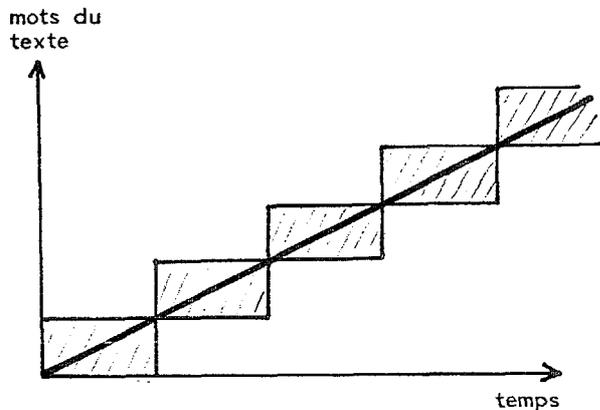


Figure 1

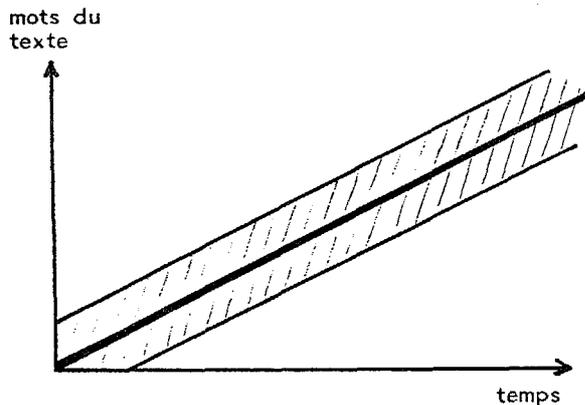


Figure 2

Une lecture du texte est une courbe  $m(t)$  dans ce plan, représentant le mot lu à l'instant  $t$ . Cette courbe doit être incluse dans le *champ de vision* puisqu'on ne peut lire que les mots visibles. On voit que la méthode par bloc impose un passage obligé à chaque fin de bloc. Cette contrainte n'est pas gênante si la vitesse moyenne de défilement est faible. Par contre, pour un défilement rapide, le risque est grand de ne pas avoir le temps de lire les derniers mots du bloc dans le cas où la vitesse de lecture varie autour de la vitesse moyenne. Au contraire, les deux autres méthodes imposent moins de contraintes sur les variations de la vitesse de lecture autour de la vitesse moyenne.

Dans le cas du décalage, il est nécessaire qu'il soit continu, car un décalage discret interviendrait trop souvent et gênerait la lisibilité.

Tous ces résultats doivent être confirmés par des expérimentations.

## CONCLUSION

Nous avons montré la possibilité matérielle d'un système de sous-titrage en temps réel pour le Français et présenté des méthodes d'apparition du texte à l'écran, en vue d'un défilement rapide.

Un système de sous-titrage pourrait être particulièrement utile pour les émissions de télévision. Néanmoins, si tous les éléments existent, il faut encore pouvoir les réunir. Cela nécessite des expérimentations, et pose des problèmes nouveaux d'ergonomie quand à la présentation de texte à l'écran pour une meilleure lisibilité.

## BIBLIOGRAPHIE

- [1] R.G. Baker, A.C. Downton, A.F. Newell 'Simultaneous speech transcription and TV captions for the deaf' Processing of visible language 2 NATO Conference series, Plenum Press, New York
- [2] I.M. Beard 'How British Palantype reporters are helping the deaf' National Shorthand Reporter, January 1983.
- [3] M.H. Block 'Captioning the Oscar' National Shorthand Reporter, August 1982.
- [4] M.H. Block, M. Okrand, 'Real-time closed-captioned television as an educational tool' American Annals for the Deaf, September 1983.
- [5] Anne-Marie Derouault, Bernard Merialdo 'TASF, a Stenotypy to French Transcription System' 7th International Conference on Pattern Recognition, Montreal 1984.
- [6] Anne-Marie Derouault, Bernard Merialdo 'Une expérience de transcription automatique de la Sténotypie en Français' Techniques et Sciences de l'Informatique, Vol 2 no 5, 1983.
- [7] Anne-Marie Derouault, Bernard Merialdo 'Recognition complexity with large vocabulary' ICASSP 84, San Diego.
- [8] J.S. Matter 'Pilot project: work with hearing-impaired students' National Shorthand Reporter, January 1983.
- [9] Ross Stuckless 'Real-time transliteration of speech into print for hearing-impaired students in regular classes' American Annals for the Deaf, September 1983.