



**HAL**  
open science

# Evaluation de l'algorithme de stylisation mélodique MOMEL et du système de codage symbolique INTSINT avec un corpus de passages en catalan

Monica Estruch

► **To cite this version:**

Monica Estruch. Evaluation de l'algorithme de stylisation mélodique MOMEL et du système de codage symbolique INTSINT avec un corpus de passages en catalan. Travaux interdisciplinaires du Laboratoire Parole et Langage, 2000, 19, pp.45-62. hal-00285558

**HAL Id: hal-00285558**

**<https://hal.science/hal-00285558>**

Submitted on 5 Jun 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.

# ÉVALUATION DE L'ALGORITHME DE STYLISATION MÉLODIQUE MOMEL ET DU SYSTÈME DE CODAGE SYMBOLIQUE INTSINT AVEC UN CORPUS DE PASSAGES EN CATALAN

Mònica Estruch<sup>1</sup>

## Résumé

*Dans cet article, nous présentons une évaluation des modules automatiques de stylisation mélodique MOMEL et de codage symbolique INTSINT avec un corpus de passages en catalan. Les résultats obtenus sont satisfaisants, le pourcentage d'erreurs étant inférieur à 10% pour MOMEL et à 20% pour INTSINT. Les erreurs sont définies pour les cas dans lesquels la courbe mélodique modélisée ou codée présente des variations perceptives par rapport à l'originale, ce qui exige une correction manuelle. Les résultats ont été comparés avec ceux obtenus pour d'autres langues dans le cadre du projet MULTEXT. Un test perceptif a permis d'étudier si les auditeurs non habitués à la voix synthétisée perçoivent des différences entre les courbes originelles et les courbes corrigées.*

*Les erreurs sont classées en fonction d'une série de variables –position des points corrigés dans la phrase, direction du déplacement et coïncidence des points déplacés avec des syllabes toniques ou atones– en vue de proposer de futures améliorations du système.*

## Abstract

*This paper presents an evaluation of the automatic modelling system MOMEL and the symbolic coding INTSINT applied to a corpus of passages in Catalan. The results obtained are satisfactory, as the error percentage is less than 10% for MOMEL and 20% for INTSINT.*

*Errors are considered to be cases in which the stylised or the coded fundamental frequency curve is perceptually different from the original one, requiring a manual correction. The results have been compared with those obtained for other languages in the framework of the MULTEXT project. A perception experiment has been carried out to test whether listeners who are not used to synthetic speech are able to perceive the differences between the original and the corrected curves.*

*Errors are classified according to different criteria –position of the modified target points, movement direction of the modified point and coincidence of the corrected points with stressed or unstressed syllables– in order to propose further system improvements.*

Mots-clés : catalan, Fo, prosodie, stylisation, MOMEL.

Keywords : catalan, Fo, prosody, stylisation, MOMEL.

---

<sup>1</sup> Grup de Fonètica  
Departament de Filologia Espanyola  
Universitat Autònoma de Barcelona  
<monica@liceu.uab.es>

## 1. Introduction

Le travail présenté dans cet article fait partie d'un projet de recherche sur la prosodie du catalan<sup>2</sup>, développé initialement dans le cadre du projet européen MULTEXT (*Multilingual Text Tools and Corpora*) (Hirst *et al.*, 1994). La validation de la stylisation automatique de la mélodie dans un corpus de passages en espagnol réalisée dans le cadre de ce projet et la constatation des possibilités d'amélioration du système ont montré le besoin d'une analyse détaillée. Il nous a donc paru utile de mener à terme cette analyse avec un corpus catalan, étant donné que cette langue n'avait pas fait partie de celles traitées dans MULTEXT<sup>3</sup>.

La première partie de cette étude concerne la validation perceptive de la stylisation automatique des passages du corpus avec l'algorithme MOMEL (*MOdelling MELody*) et de l'étiquetage avec INTSINT (*INternational Transcription System for INTonation*), modules développés au laboratoire Parole et Langage d'Aix-en-Provence (Hirst & Di Cristo, 1998).

Cet article a donc comme objectif principal la présentation des résultats obtenus concernant la validation du module MOMEL (section 3) et l'étiquetage avec INTSINT (section 4) dans un corpus catalan. Ils sont précédés d'un bref aperçu de la méthodologie utilisée (section 2) pour la constitution du corpus, la sélection des locuteurs, l'enregistrement et l'analyse du corpus. Les conclusions de l'analyse sont présentées à la fin de l'article.

## 2. Protocole expérimental

### 2.1. Corpus

Le corpus de ce travail est constitué de 40 passages de 5 phrases chacun ; chaque passage est construit de façon à ce qu'il soit sémantiquement cohérent. Pour faciliter la comparaison interlangues, ces passages sont une traduction et adaptation au catalan du corpus multilingue Eurom.1 développé dans le projet ESPRIT SAM (*Multi-lingual Speech Input/Output Assessment*,

---

2. Les résultats de la première phase du projet sont recueillis dans le mémoire «Analyse mélodique et codification symbolique d'un corpus de passages en catalan», réalisée dans le cadre du programme de troisième cycle «Linguistique: traitement informatique du langage» à l'Universitat Autònoma de Barcelona (Estruch, 1999). Cette étude a été dirigée par le professeur Joaquim Llisterra et soutenue en janvier 2000 devant un jury constitué par les professeurs Joan A. Argenter, Lorraine Baqué et Juan M. Garrido, que je remercie de tous ses précieux commentaires. Le travail a été financé par une bourse de formation à la recherche accordée par le ministère d'Éducation et de Culture espagnol (1997-2000). Une bourse de voyage du même organisme m'a permis de réaliser un stage à Aix-en-Provence, où j'ai profité du généreux accueil et des conseils de Daniel Hirst et de Robert Espesser.

3. Jusqu'à présent, nous avons travaillé, d'un côté, à la validation perceptive de l'algorithme de stylisation MOMEL et du système de codification symbolique INTSINT, présentés dans cet article, et d'un autre côté, à l'adéquation de ce système à la description des phénomènes prosodiques du catalan : plus précisément, nous avons étudié le rapport entre la courbe de fréquence fondamentale et les pauses –réajustements fréquentiels, mouvements avant pause– ; le rapport entre les sommets de F<sub>0</sub> et les syllabes toniques ou atones; et l'évolution, dans les passages, des points Top, Mid et Bottom du système INTSINT par rapport à la déclinaison. Les résultats sont présentés dans Estruch (1999).

*Methodology and Standardisation*) (Chan *et al.*, 1995) et utilisé dans MULTEXT. Nous avons donc 200 phrases qui ont été enregistrées par deux locuteurs différents, soit un total de 400 phrases. La durée totale de l'enregistrement est de 24 minutes et 25 secondes.

## 2.2. Locuteurs

Les deux locuteurs qui ont réalisé les enregistrements sont un locuteur masculin (JV) et une locutrice féminine (LP) de langue maternelle catalane, âgés respectivement de 27 et 25 ans et possédant une formation universitaire.

## 2.3. Enregistrement

Les enregistrements ont eu lieu dans le studio insonorisé du Laboratoire de Phonétique de l'Universitat Autònoma de Barcelona.

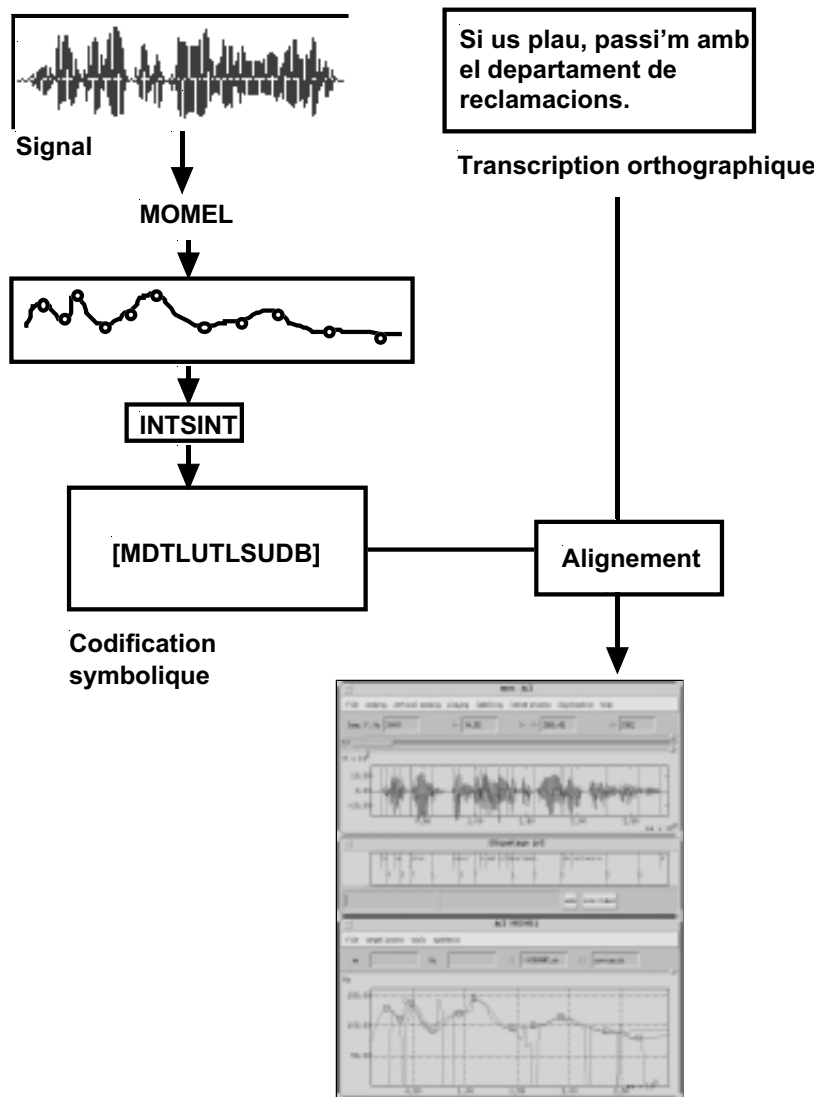
Le signal a été numérisé et analysé avec le logiciel MES (*Motif Environment for Speech*) développé au laboratoire Parole et Langage d'Aix-en-Provence (Espesser, 1996). Ce logiciel possède un module de stylisation automatique, MOMEL, remplaçant la courbe de  $F_0$  par une courbe lisse et continue qui représente l'évolution de la  $F_0$  moyennant une concaténation de points-cibles (Hirst & Espesser, 1993). Il existe aussi un module d'étiquetage automatique de ces points-cibles, INTSINT (Hirst *et al.*, 2000) qui permet de décrire la courbe de  $F_0$  à l'aide de 8 symboles (Top, Mid et Bottom, tons absolus qui décrivent la tessiture fréquentielle du locuteur, et Higher, Lower, Upstep, Downstep et Same, des tons relatifs qui tiennent compte du point-cible précédent).

Le processus suivi est celui de la figure 1.

## 3. Validation du système de stylisation MOMEL

La première partie de l'analyse a eu pour but la validation de la stylisation automatique réalisée avec MOMEL. Tel qu'il a été déjà signalé, ce module simplifie la courbe de  $F_0$  originelle en une courbe lisse et continue avec des points-cibles.

Cet algorithme automatique a été utilisé pour plusieurs langues (dans le cadre du projet MULTEXT, voir par exemple (Llisterri, 1996; Astesano *et al.*, 1997; Campione & Véronis, 1998; Véronis & Campione, 1998) : en général, les résultats obtenus sont satisfaisants, mais il reste cependant quelques erreurs dans la stylisation des énoncés. Ceci exige une correction manuelle basée sur une comparaison perceptive de la courbe originale et de la courbe modélisée.

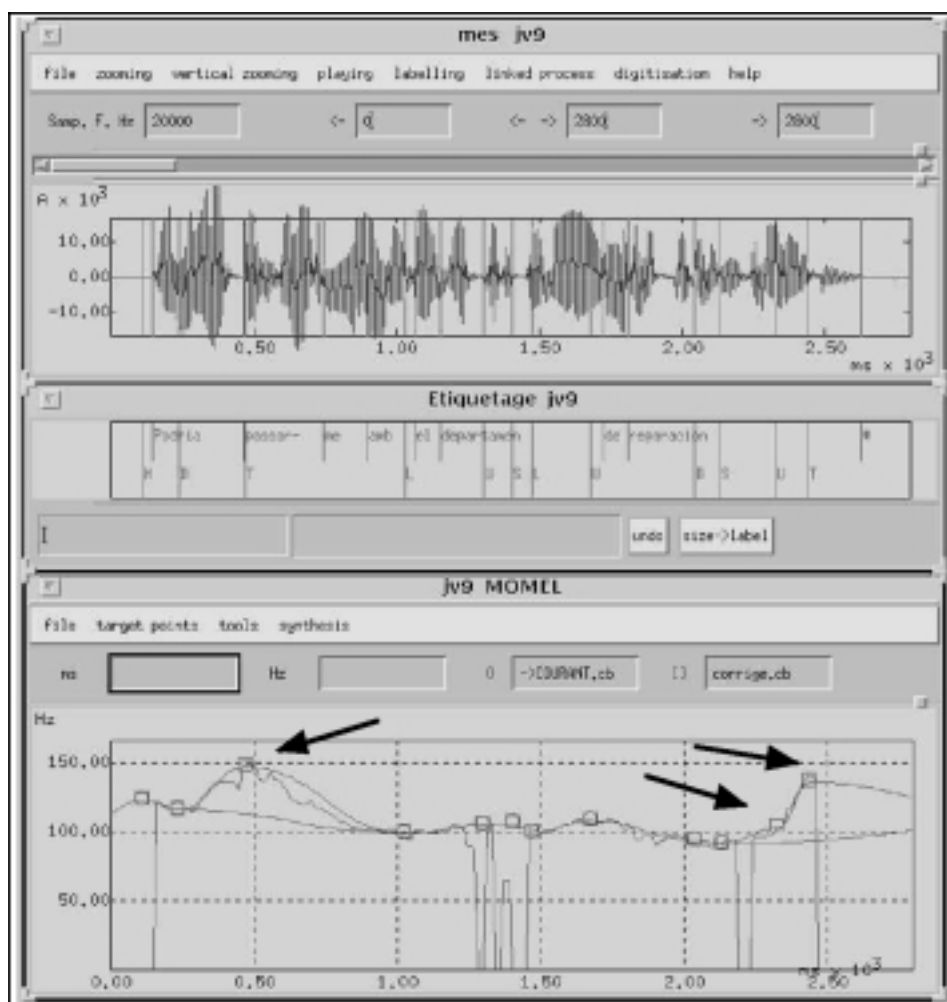


**Figure 1.**

*Schéma du processus de traitement du signal vocal avec le logiciel MES appliqué à un exemple de phrase du corpus catalan (« Passez-moi les réclamations, s'il vous plaît. »)*

En supprimant, ajoutant ou déplaçant des points-cibles on doit arriver à une équivalence perceptive entre les réalisations acoustiques des deux courbes. Dans cette section nous présentons la systématisation des modifications nécessaires pour obtenir, à partir des courbes stylisées générées par MOMEL, une équivalence perceptive avec les courbes originales de notre corpus en catalan.

On peut observer un exemple des différences entre les deux courbes figure 2.



**Figure 2.**

*Phrase produite pour le locuteur masculin; dans la partie inférieure, on observe la courbe originelle (points ronds) et la courbe corrigée avec les points-cibles qu'on a ajoutés (points carrés) indiqués avec une flèche.*

Les résultats obtenus après la validation sont ceux présentés dans le tableau 1:

	Locutrice LP		Locuteur JV		Total
	Nombre de cas	Pourcentage	Nombre de cas	Pourcentage	
Points ajoutés	213	9,13%	161	6,23%	4.913
Points déplacés	30	1,28%	22	0,85%	374
Points courbes originales	2.331		2.582		52

**Tableau 1.**

*Nombre total et pourcentage de points-cibles corrigés (ajoutés), de points déplacés, de points des courbes originales et de points des courbes après la correction, pour les deux locuteurs.<sup>4</sup>*

Afin d'établir un classement plus précis des erreurs détectées, on a choisi le critère de la présence ou absence de pause pour décrire la localisation des points qui ont dû être modifiés. On peut trouver, donc, trois positions différentes:

- après pause (début absolu de phrase ou après une pause intérieure de phrase)
- avant pause (fin de phrase –avec un contour ascendant ou descendant– ou avant une pause intérieure de phrase)
- position intérieure de phrase.

### 3.1. Points-cibles ajoutés

On peut observer dans le tableau 1 qu'il y a 213 cas de points ajoutés pour la locutrice LP, ce qui représente 9,13% du total des points-cibles existants, et 161 cas pour le locuteur JV, 6,23% du total. Le tableau 2 montre ces points classés en fonction des différentes positions dans la phrase.

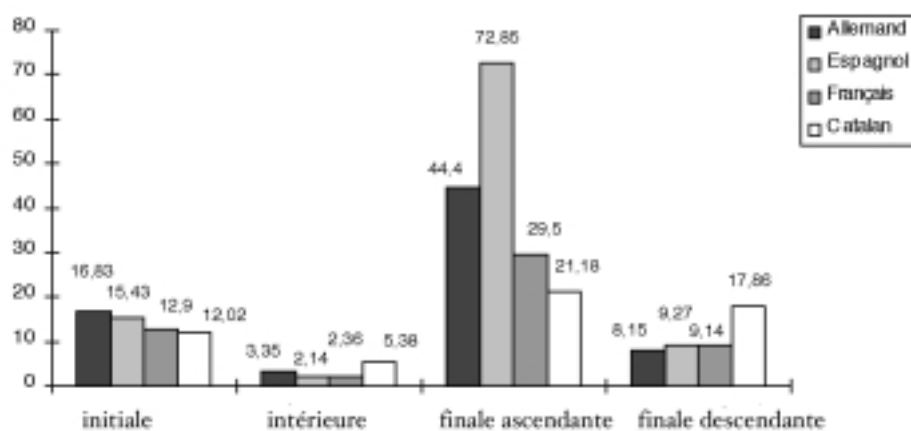
Position	Locutrice LP		Locuteur JV	
	Nombre de cas	Pourcentage	Nombre de cas	Pourcentage
Après pause	54/313	17,25%	18/265	6,79%
Intérieure	98/1.705	5,74%	103/2.052	5,02%
Avant pause contour ascendant	26/68	38,24%	4/97	4,12%
Avant pause contour descendant	35/245	14,29%	36/168	21,43%
TOTAL	213/2.331	9,13%	161/2.582	6,23%

**Tableau 2.**

*Points-cibles ajoutés dans tout le corpus pour les deux locuteurs: nombre de cas et pourcentages en fonction de la position dans la phrase.*

<sup>4</sup> On peut observer qu'il n'y a aucun cas de suppression de points-cibles de la courbe stylisée.

Comme nous l'indiquions dans l'introduction, cette validation de MOMEL avait déjà été effectuée pour d'autres langues dans le cadre de MULTEXT. Il est donc possible de comparer les résultats obtenus pour le français, l'allemand et l'espagnol (Llisterri, 1996) avec ceux du catalan.



**Figure 3.**  
*Pourcentage de points ajoutés en allemand, espagnol, français et catalan en fonction de la position des points dans la phrase.*

On peut observer dans la figure 3 une variabilité assez marquée dans les pourcentages, en fonction de la position du point-cible, notamment en position finale ascendante. Cependant, il faut remarquer qu'il n'est pas possible d'identifier une langue dans laquelle le nombre global d'erreurs soit nettement supérieur ou inférieur à celui des autres.

### 3.2. Points-cibles déplacés

Dans certains cas, il faut non pas ajouter des nouveaux points-cibles, mais effectuer un déplacement de ceux que le système a détectés pour obtenir une courbe modélisée perceptivement équivalente à l'originale. Dans ce cas, nous avons déplacé 30 points pour la locutrice LP, ce qui représente 1,28% du total des points-cibles existants, et 22 pour le locuteur JV, soit 0,85% du total. Dans le tableau 3, on observe la distribution de ces points en fonction des positions dans la phrase.

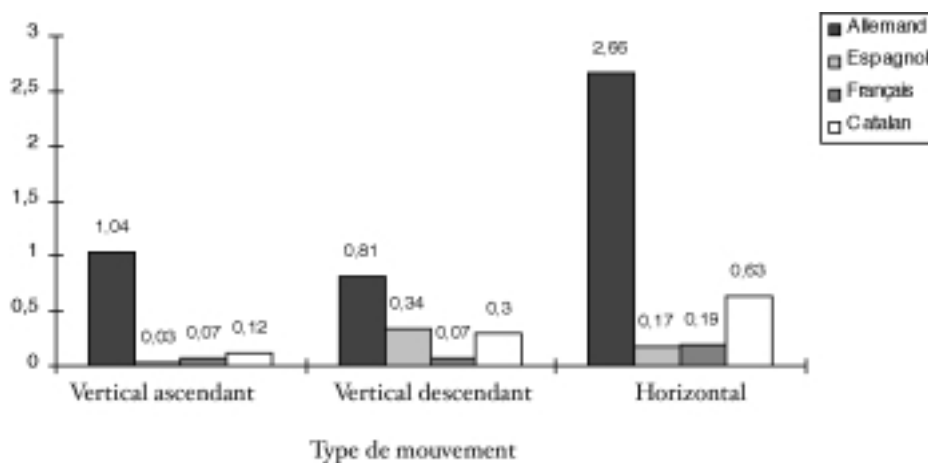


Position	Locutrice LP		Locuteur JV	
	Nombre de cas	Pourcentage	Nombre de cas	Pourcentage
Après pause	4/313	1,27%	2/265	0,75%
Intérieure	19/1.705	1,11%	14/2.052	0,68%
Avant pause contour ascendant	1/68	1,47%	1/97	1,03%
Avant pause contour descendant	6/245	2,44%	5/168	2,97%
TOTAL	30/2.331	1,28%	22/2.582	0,85%

**Tableau 3.**

*Points-cibles déplacés dans tout le corpus pour les deux locuteurs:  
nombre de cas et pourcentages en fonction de la position dans la phrase.*

La figure 4 présente les résultats du catalan comparés avec ceux obtenus pour le français, l'allemand et l'espagnol. Nous avons classé les points en fonction du type de mouvement de déplacement, pour unifier les critères avec ceux utilisés dans MULTTEXT.



**Figure 4.**

*Pourcentage de points ajoutés en allemand, espagnol,  
français et catalan en fonction du type de mouvement de déplacement.*

On constate dans la figure 4 qu'il y a peu de variation entre les pourcentages du français, de l'espagnol et du catalan, contrairement au cas de l'allemand, où le nombre de points déplacés est beaucoup plus élevé. Ceci est dû au fait que, pour la validation de cette langue, nous avons eu tendance à déplacer beaucoup plus de points au lieu d'en ajouter de nouveaux, critère qui n'a pas été suivi dans les autres langues<sup>5</sup>.

### 3.3. Conclusions sur la validation de la stylisation

On peut systématiser les résultats obtenus de la manière suivante:

- pour la locutrice féminine LP, la plupart des erreurs se trouvent dans la position avant pause en contour ascendant (il faut remarquer que MOMEL ne détecte pas le point-cible final avant pause quand la fréquence est très élevée) et après pause.
- pour le locuteur masculin JV, les erreurs apparaissent généralement avant pause en contour descendant (le système ne détecte pas le point-cible final avant pause si la fréquence est très basse) et après pause.

On peut donc affirmer que les erreurs se situent majoritairement dans les positions avant et après pause. Il en découle que l'amélioration du traitement des pauses dans MOMEL permettrait d'obtenir une meilleure stylisation.

Dans les positions avant et après pause, les corrections dans la phase de validation consistent surtout à ajouter de nouveaux points-cibles; dans les positions intérieures de phrase, il faut, généralement, déplacer les points déjà existants.

### 3.4. Test perceptif

Une question que nous nous posions au début de l'étude était la possibilité de réaliser un test perceptif permettant d'évaluer si les changements de points-cibles effectués sur les courbes modélisées étaient également perçus par des sujets non habitués à la voix synthétisée. Les travaux publiés sur la validation de MOMEL n'ont pas, à notre connaissance, adopté cette approche, qui permettrait de vérifier si les modifications introduites dans la courbe corrigée sont ou non perçues par d'autres sujets autres que le chercheur qui réalise la validation.

Nous avons donc préparé un test avec des paires de phrases (ou parties cohérentes de phrases) extraites du corpus et présenté 20 paires de phrases différentes (la version originelle et la corrigée dans les deux ordres possibles) et 20 paires de phrases identiques (paires de contrôle pour vérifier que les auditeurs ne répondaient pas de manière aléatoire) pour les deux locuteurs. Nous avons obtenu 80 paires de phrases, que nous avons enregistrées sur une cassette puis fait écouter, dans des conditions équivalentes, à 15 auditeurs de langue maternelle catalane non

---

<sup>5</sup> Les critères suivis pour chaque langue sont décrits par Llisterra (1996).

habitué à la voix synthétisée. Ces auditeurs devaient écouter les paires de phrases, décider si elles leur semblaient égales ou différentes du point de vue de la mélodie et signaler leur choix sur une feuille de réponses. Pour la discussion des résultats, nous considérons qu'une réponse est «correcte» lorsque l'auditeur identifie comme égales des paires avec les mêmes stimuli, et comme différentes des paires de stimuli différents.

Les résultats obtenus sont présentés dans le tableau 4 :

Locuteur	O-O	O-C	C-O	C-C	TOTAL
LP	96,60%	76%	79,30%	90,60%	85,60%
JV	92,60%	58%	42%	93,30%	71,50%

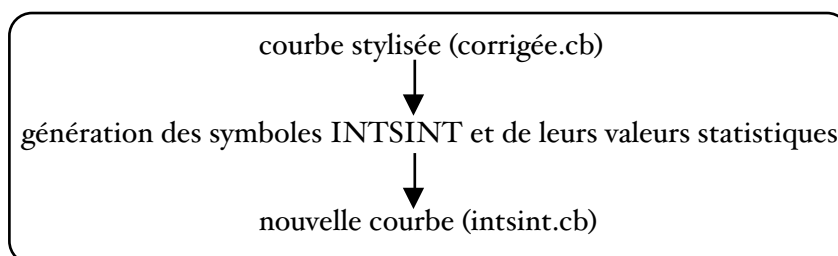
**Tableau 4.**

*Pourcentage de réponses correctes obtenues pour les deux locuteurs dans le test perceptif. 'O' indique la courbe originelle et 'C' la courbe corrigée.*

Les chiffres obtenus montrent d'abord que les auditeurs ne répondent pas aléatoirement : pour les deux locuteurs, les résultats des paires de contrôle O-O et C-C sont supérieurs à 90%. Dans les paires formées par une version originelle et une version corrigée de la courbe, les résultats sont bien différenciés pour les deux locuteurs: dans le cas de la locutrice LP, les réponses correctes dépassent 75%; en revanche, en ce qui concerne le locuteur JV, les auditeurs n'ont pas réussi de la même façon à percevoir si les deux phrases de la paire étaient identiques ou non, avec des résultats autour de 50%. Ceci semble indiquer une certaine dépendance du locuteur, dont il faut peut-être tenir compte au moment de valider la stylisation.

#### 4. Validation du système d'étiquetage symbolique INTSINT

La deuxième partie de cette analyse concerne la validation du module d'étiquetage symbolique INTSINT. Comme on l'a déjà signalé, ce système assigne un symbole à chacun des points-cibles détectés dans la courbe modélisée. On peut schématiser le processus comme suit (figure 5) :



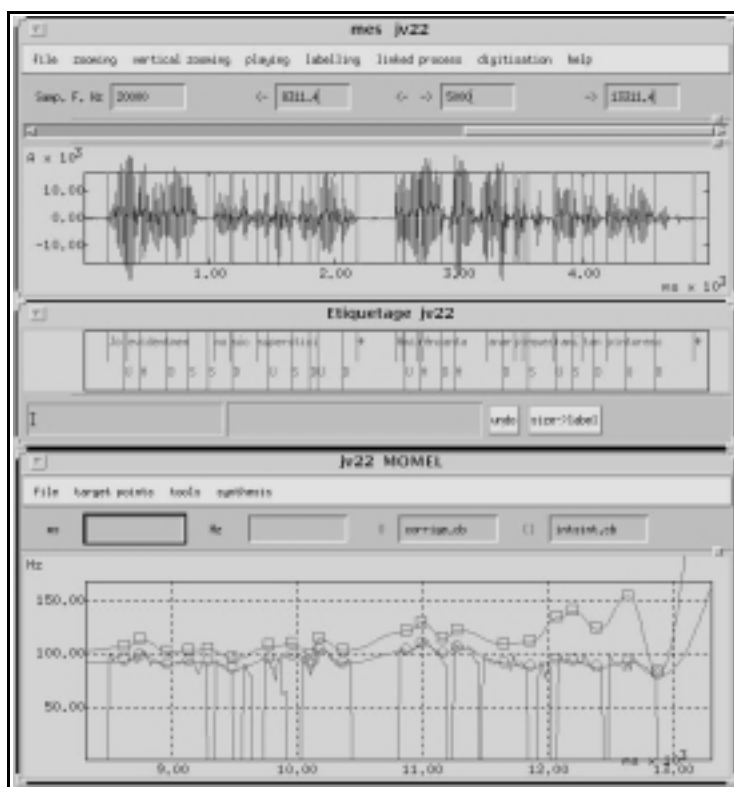
**Figure 5.**

*Schéma du processus de génération de la nouvelle courbe Intsint à partir de la courbe modélisée corrigée.*

L'objectif final de la génération de la courbe INTSINT est de vérifier s'il est possible de générer une courbe perceptivement équivalente à l'originale uniquement à partir des symboles et de ses valeurs statistiques<sup>6</sup>.

Comme on peut l'observer dans la figure 5, les symboles d'INTSINT sont créés à partir de la courbe stylisée corrigée: pour cette raison, il est essentiel que la courbe de MOMEL ne contienne pas de points-cibles perceptivement différents de l'originale; sinon, ces erreurs pourraient se reproduire aussi dans la courbe INTSINT.

Néanmoins, il arrive parfois que la nouvelle courbe générée à partir des symboles INTSINT ne soit pas adéquate. C'est le cas de l'exemple présenté dans la figure 6 :



**Figure 6.**

*Phrase produite par le locuteur masculin;  
dans la partie inférieure, on observe la courbe corrigée (points ronds)  
et la nouvelle courbe à partir des symboles INTSINT avec des erreurs de génération (points carrés).*

6. Une des phases de codification de la courbe de  $F_0$  avec le système INTSINT inclut le calcul des valeurs statistiques de chacune des 8 catégories de symboles: la moyenne dans le cas des tons absolus, et une représentation des coefficients de régression en supposant que le point est modélisé par une fonction de type  $P_i = a \cdot p_i - 1 + b$ .

La validation menée à terme dans ce cas a consisté à comparer perceptivement la courbe corrigée de MOMEL (déjà équivalente à l'originale) avec celle générée à partir des symboles d'INTSINT et des valeurs statistiques : il a fallu déplacer des points-cibles pour obtenir l'équivalence perceptive entre les deux courbes.

#### 4.1. Points-cibles déplacés

Les erreurs détectées sont classées, d'abord, en fonction de tous les symboles apparaissant dans le corpus (tableau 5); le tableau 6 présente un classement en fonction des symboles appartenant à chaque catégorie.

Symboles	Locutrice LP		Locuteur JV	
	Nombre de cas	Pourcentage	Nombre de cas	Pourcentage
T	20	4,15%	7	1,84%
M	5	1,04%	0	0%
B	8	1,66%	1	0,26%
H	78	16,18%	47	12,34%
L	16	3,32%	24	6,30%
U	148	30,70%	119	31,23%
D	157	32,58%	98	25,72%
S	50	10,37%	85	22,31%
TOTAL	482	100%	381	100%

**Tableau 5.**

*Nombre de cas et pourcentage (sur le total du corpus) des points-cibles déplacés sur la courbe intsint.cb pour chacun des locuteurs.*

Dans le tableau 5 on peut observer le nombre de points-cibles qu'on a déplacés de chaque catégorie de symboles pour les deux locuteurs. On note que le total est plus important dans le cas de la locutrice LP (482) que pour le locuteur JV (381). Si l'on considère les différentes catégories de symboles, celles qui présentent un pourcentage plus élevé d'erreurs sont D et U, et ce pour les deux locuteurs. D'autre part, les catégories qui correspondent aux tons absolus sont celles qui ont les pourcentages d'erreurs les plus faibles.

Remarquons que ces résultats ont été calculés par rapport à tous les points déplacés du corpus, c'est-à-dire, qu'on a déplacé par exemple 20 points-cibles codés T de l'ensemble des 482 points déplacés pour la locutrice LP (soit 4,15% dans le tableau).

Une autre approche pour calculer le nombre de points déplacés pour chaque symbole est de ne considérer que le total de points-cibles codés avec ce symbole; par exemple, sur l'ensemble

des points codés comme T pour un locuteur, on calcule combien d'entre eux ont été déplacés (7/120 dans le cas du locuteur masculin). Ces résultats sont présentés dans le tableau 6.

Symboles	Locutrice LP		Locuteur JV	
	Nombre de cas	Pourcentage	Nombre de cas	Pourcentage
T	20/143	13,99%	7/120	5,83%
M	mai-97	5,15%	0/87	0%
B	8/214	3,74%	1/113	0,88%
H	78/302	25,83%	47/337	13,95%
L	16/243	6,58%	24/372	6,45%
U	148/556	26,62%	119/616	19,32%
D	157/736	21,33%	98/714	13,72%
S	50/253	19,76%	85/384	22,13%
TOTAL	482/2.544	18,95%	381/2.743	13,89%

**Tableau 6.**  
*Nombre de points déplacés codés avec le système INTSINT  
sur le total de symboles, et pourcentage correspondant, pour les deux locuteurs.*

Dans le tableau 6, on peut observer que le pourcentage total d'erreurs – points-cibles qu'on a dû déplacer pour rendre la courbe résultante perceptivement semblable à la courbe modélisée corrigée – est supérieur dans le cas de la locutrice LP (18,95%) au cas du locuteur JV (13,89%). On remarque aussi dans le tableau que les catégories dont le pourcentage est le plus élevé sont, pour les deux locuteurs, D et U: autour de 20-25% dans le cas de la locutrice féminine et de 15-20% dans le cas du locuteur masculin. S et H sont également des catégories avec un pourcentage d'erreurs très élevé.

#### 4.2. Distribution des points déplacés

Nous avons analysé une série de variables pour essayer de décrire d'une façon plus précise les erreurs détectées dans la nouvelle courbe générée à partir des symboles INTSINT (intsint.cb). Les variables considérées sont les suivantes :

- nombre de cas de points déplacés en direction ascendante et descendante et valeur moyenne en Hz du déplacement de la fréquence fondamentale ;
- points-cibles déplacés en fonction de la position dans la phrase (avant pause, après pause ou à l'intérieur de la phrase) ;
- nombre de points déplacés coïncidant avec des syllabes toniques ou atones.

#### 4.2.1. Direction du déplacement

En ce qui concerne le nombre total de points à corriger pour les deux locuteurs, le pourcentage correspondant à des points déplacés en direction descendante (points situés trop hauts dans l'axe fréquentiel) est un peu supérieur au pourcentage de points dont on a augmenté la fréquence (points situés trop bas dans l'axe fréquentiel): 53% vs. 46% en total, si l'on prend les deux locuteurs ensemble.

La valeur moyenne en Hz des déplacements dans les deux directions est clairement supérieure dans le cas de la locutrice féminine (entre 19 et 35 Hz pour les déplacements descendants et entre 27 et 37 Hz pour les déplacements ascendants); le locuteur masculin présente des moyennes entre 11 et 14 Hz dans les déplacements ascendants et descendants. Ces deux tendances exposées, il faut les lier au fait que le rang fréquentiel de la locutrice féminine est beaucoup plus large (275 Hz) que celui du locuteur masculin (84 Hz).

Les résultats détaillés se trouvent dans le tableau 7:

Symboles	Locutrice LP				Locuteur JV			
	Ascendant		Descendant		Ascendant		Descendant	
	Nombre de cas	Moyenne en Hz	Nombre de cas	Moyenne en Hz	Nombre de cas	Moyenne en Hz	Nombre de cas	Moyenne en Hz
T	11	29	9	29,6	4	11,5	3	11
M	4	33	1	26	0	-	0	-
B	7	33,5	1	19	1	14	0	-
H	46	32,5	32	32,7	26	12,1	21	13,9
L	12	27,2	4	34,4	11	12,5	13	11,3
U	65	36,3	83	32,9	49	12,4	70	13,4
D	68	34	89	29,7	47	11,7	51	11,9
S	13	31,7	37	30,2	40	12,2	45	12,2
TOTAL	226		256		178		203	

**Tableau 7.**

*Nombre de cas et moyenne en Hz du déplacement des points-cibles en fonction de la direction de ce déplacement, pour les deux locuteurs.*

#### 4.2.2. Position des points déplacés

Comme on peut l'observer dans le tableau 8, le nombre de points corrigés en position intérieure de phrase est supérieur au nombre de points corrigés dans les deux autres positions (avant et après pause) pour les deux locuteurs. On peut expliquer cette tendance en tenant compte du fait que le nombre de points se trouvant dans cette position dans le corpus (et son pourcentage) est supérieur au nombre de points placés avant ou après une pause.

Symboles	Locutrice LP						Locuteur JV					
	Après pause		Avant pause		Intérieur		Après pause		Avant pause		Intérieur	
	cas	%	cas	%	cas	%	cas	%	cas	%	cas	%
T	2	3,7	2	3,57	16	4,3	2	9,09	1	2,04	4	1,29
M	3	5,55	0	-	2	0,54	0	-	0	-	0	-
B	0	-	3	5,36	5	1,34	0	-	1	2,04	0	-
H	7	12,96	2	3,57	69	18,55	1	4,54	5	10,2	41	13,22
L	0	-	3	5,36	13	3,49	0	-	5	10,2	19	6,13
U	38	70,37	18	32,14	92	24,73	10	45,45	14	28,57	95	30,64
D	3	5,55	16	28,59	138	37,1	6	27,27	12	24,49	80	25,81
S	1	1,85	12	21,43	37	9,95	3	13,64	11	22,45	71	22,9
TOTAL	54	11,2	56	11,62	372	77,18	22	5,77	49	12,86	310	81,15

**Tableau 8.**  
*Nombre de cas des points-cibles déplacés  
en fonction de la position, pour les deux locuteurs.*

#### 4.2.3. Coïncidence des points déplacés avec syllabes toniques ou atones

En se référant au tableau 9, on constate que le nombre total de points corrigés qui coïncident avec des syllabes toniques est légèrement inférieur à celui des points qui coïncident avec des syllabes atones, et ce pour les deux locuteurs.

Symboles	Locutrice LP				Locuteur JV			
	Toniques		Atones		Toniques		Atones	
	cas	%	cas	%	cas	%	cas	%
T	8	3,74	12	4,48	2	1,09	5	2,54
M	3	1,4	2	0,75	0	-	0	-
B	6	2,8	2	0,75	1	0,54	0	-
H	30	14,02	48	17,91	16	8,69	31	15,74
L	10	4,67	6	2,24	9	4,89	15	7,61
U	63	29,44	85	31,72	57	30,98	62	31,47
D	76	35,51	81	30,22	52	28,26	46	23,35
S	18	8,41	32	11,94	47	25,54	38	19,29
TOTAL	214	44,4	268	55,6	184	48,29	197	51,71

**Tableau 9.**  
*Nombre de cas des points-cibles déplacés qui coïncident avec des syllabes toniques ou atones.*



## 5. Conclusion

L'analyse du processus de validation des modules de stylisation (MOMEL) et d'étiquetage (INTSINT) appliqué à un corpus catalan permet d'arriver à trois conclusions principales :

- Le système de stylisation automatique MOMEL présente peu d'erreurs : un maximum de 10% pour l'ensemble du corpus. Ce pourcentage correspond essentiellement à des points-cibles situés avant et après pause. On peut donc penser qu'une amélioration du traitement des pauses diminuerait d'une manière très significative le pourcentage total d'erreurs.
- La réalisation d'un test perceptif consistant à faire écouter des paires de phrases formées par la courbe originelle et la courbe corrigée à des auditeurs non habitués à la voix synthétisée, à qui on a demandé de répondre si les deux phrases leur ont semblé égales ou différentes du point de vue de la mélodie, nous a permis de vérifier que la différence est perçue à des taux supérieurs à 75% pour la locutrice féminine et proches de 50% pour le locuteur masculin; on peut donc penser que les auditeurs perçoivent, comme le chercheur qui réalise la validation, les changements de points-cibles effectués dans les courbes modélisées.
- Le système d'étiquetage automatique INTSINT présente moins de 20% d'erreurs pour les deux locuteurs dans la génération d'une nouvelle courbe à partir du codage symbolique. Dans ce cas, il est plus difficile de systématiser dans quels contextes le système produit des erreurs : d'un côté, il y a un nombre important de points dont on doit abaisser la  $F_0$ , c'est-à-dire, des points-cibles que le système tend à situer à une fréquence trop élevée; d'un autre côté, la majorité des points déplacés se trouvent en position intérieure de phrase; finalement, pour les deux locuteurs étudiés, il y a un nombre légèrement supérieur de points corrigés qui coïncident avec des syllabes atones.

Ces résultats montrent, d'une part, qu'il est possible d'utiliser MOMEL et INTSINT pour des langues différentes de celles considérées jusqu'à présent. D'autre part, ils suggèrent quelques possibles améliorations d'un système dont la performance est cependant déjà très satisfaisante. Finalement, le travail présenté ici ouvre la voie non seulement à une analyse de la prosodie du catalan fondée sur le codage INTSINT déjà initiée, mais aussi à de futures comparaisons inter-linguistiques.

## 6. Bibliographie

- ASTESANO, C.; ESPESSER, R.; HIRST, D.; LLISTERRI, J. (1997). Stylisation automatique de la fréquence fondamentale: une évaluation multilingue. In *Actes du 4ème Congrès Français d'Acoustique*. Marseille, avril 1997, p. 441-444.
- CAMPIONE, E.; VERONIS, J. (1998). A statistical study of pitch target points in five languages. In *Proceedings of the 5th International Conference on Spoken Language Processing*. Sidney, Australia, nov. 30 - dec. 4 1998, vol.4, p. 1391-1394.
- CHAN, D.; FOURCIN, A.; GIBBON, D.; GRANSTRÖM, B.; HUCKVALE, M.; KOKKINAKIS, G.; KVALE, K.; LAMEL, L.; LINDBERG, B.; MORENO, A.; MOUROPOULOS, J.; SENIA, F.; TRANCOSO, I.; VELD, C.; ZEILIGER, J. (1995). EUROM - A Spoken Language Resource for the EU. In J.M. Pardo *et al.* (eds.) *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology*. Madrid, Spain, 18-21 sept. 1995, vol. 1, p. 867-870.
- ESPESSER, R. (1996). MES : un environnement de traitement du signal, *Actes, 21èmes Journées d'Etude sur la Parole*, Avignon, France, 10-14 juin 1996, p. 447.
- ESTRUCH, M. (1999). *Anàlisi melòdica i codificació simbòlica d'un corpus de paràgrafs en català*. Mémoire de recherche du programme de doctorat «Linguistique: traitement informatique du langage». Departament de Filologia Espanyola, Universitat Autònoma de Barcelona.
- HIRST, D.J.; ESPESSER, R. (1993). Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix*, 15, p. 75-85.
- HIRST, D.J.; IDE, N.; VERONIS, J. (1994). Coding fundamental frequency patterns for multilingual synthesis with INTSINT in the MULTTEXT project. In *Conference Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis*. Sept. 12-15, 1994. Mohonk Mountain House, New Paltz, New York, USA, p. 77-80.
- HIRST, D.J.; DI CRISTO, A. (1998). A Survey of Intonation Systems. In D. Hirst & A. Di Cristo (eds) *Intonation Systems. A Survey of Twenty Languages*. Cambridge: Cambridge University Press, p. 1-44.
- HIRST, D.J.; DI CRISTO, A.; ESPESSER, R. (2000). Levels of representation and levels of analysis for the description of intonation systems. In M. Horne (ed.) *Prosody: Theory and Experiments*. Dordrecht, Kluwer.
- LLISTERRI, J. (ed) (1996). Prosody Tools Efficiency and Failures. *WP 4 Corpus*. T4.6 Speech Markup and Validation. Deliverable 4.5.2. Final version, 15 oct. 1996. LRE Project 62-050 MULTTEXT.
- VERONIS, J.; CAMPIONE, E. (1998). Towards a reversible symbolic coding of intonation. In *Proceedings of the 5th International Conference on Spoken Language Processing*. Sidney (Australia), nov. 30- dec; 4 1998, vol.7, p. 2899-2902.



