

Entends-tu mes attitudes ? Perception de la prosodie des affects sociaux en chinois Mandarin

Yan Lu¹ Véronique Aubergé¹ Albert Rilliard²

(1) GIPSA Lab, Université de Grenoble

(2) LIMSI-CNRS, BP133, Orsay

{yan.lu, veronique.auberge}@gipsa-lab.grenoble-inp.fr,
albert.rilliard@limsi.fr

RESUME

Les affects sociaux sont, au contraire des émotions, des actes de parole volontairement contrôlés et construits socio-culturellement. Ce travail examine la perception d'affects sociaux chinois par des sujets natifs, avec l'enseignement de la prosodie attitudinale comme finalité. Un corpus de parole a été enregistré, comportant des variations de longueur, de placement de tons et de structure syntaxique des énoncés. Tous les énoncés ont été produits avec 19 affects sociaux. Un test perceptif montre que les affects sociaux sont globalement bien reconnus : les expressions de « déclaration » et de « déception » recevant les meilleurs scores ; la « confiance » et l'« ironie » les moins bons. Tous les affects sociaux testés s'organisent de façon cohérente en sept classes conceptuelles.

ABSTRACT

Do you hear my attitudes? Perception of Mandarin Chinese social affects' prosody

Social affects are, contrary to emotions, speech acts voluntarily controlled and socio-culturally built. This work examines the perception of Chinese social affects by natives, in the aim of attitudinal prosody teaching. A speech corpus was designed, with variation of length, tone location and syntactic structures of utterances, and produced with 19 social affects. The perception test reveals that social affects were globally recognized, the expressions of "declaration" and "disappointment" received the best scores, and "confidence" and "irony" the lowest. The social affects were organized into seven conceptually coherent clusters.

MOTS-CLES : perception de la prosodie, attitudes, affects sociaux, chinois mandarin

KEYWORDS: perception of prosody, attitudes, social affects, Mandarin Chinese

1 Introduction

Les affects exprimés durant une communication interactive impliquent deux niveaux différents de processus cognitif (Aubergé, 2002) : l'expression involontairement contrôlée d'affects (les *émotions*), et l'expression intentionnellement contrôlée et transmise au travers de la prosodie audio-visuelle (les *affects sociaux* ou *attitudes*). En opposition aux émotions, les affects sociaux sont étroitement liés au langage et s'insèrent dans une culture donnée au sein de laquelle ils sont acquis durant l'enfance. Ils constituent une partie indispensable de la construction de l'interaction verbale, de la communication. Différents facteurs influents sur l'organisation des affects sociaux ont été proposés (Wichmann, 2000 ; de Moraes *et al.*, 2010), et nous proposons ici un classement assez proche :

- L'attitude exprime l'intention ou l'opinion du locuteur sur ce qu'il dit : son

engagement dans l'acte de langage (au sens de Daneš, 1994). De la même manière, le fait qu'il n'exprime (ou ne veuille, ne doive, ne puisse exprimer) aucune attitude est également considéré comme une attitude (par ex. une déclaration ou une question simple).

- Des attitudes particulières sont liées aux caractéristiques de la relation sociale (comme la hiérarchie sociale, la puissance relative) entre les locuteurs impliqués dans l'interaction (par ex. la politesse, l'autorité).
- Certaines attitudes sont liées au contexte socioculturel de l'interaction : plus spécialement pour l'intimité, le langage maternel ou la séduction.

À la suite des travaux de Martins-Baltar (1977) ou Fónagy (1991), des chercheurs ont entrepris des études sur la prosodie attitudinale dans différentes langues (Fujisaki & Hirose, 1993 ; Mejvaldova, 2000 ; Diaferia, 2002 ; Mac *et al.*, 2010 ; Gu *et al.*, 2011), et certains d'entre eux avaient plus particulièrement comme finalité l'enseignement de la prosodie attitudinale (Shochi *et al.*, 2010) ou la comparaison interculturelle des attitudes (Shochi *et al.*, 2009).

Le présent travail constitue la première étape du développement d'une méthode didactique de l'enseignement de la prosodie attitudinale française aux apprenants chinois. Cette étude analysera la reconnaissance des affects sociaux chinois par des natifs et les proximités entre les 19 affects sociaux. Nous examinerons également la distance entre perception acoustique et compréhension des concepts des affects sociaux.

Notons que la relation entre langue et culture varie à la fois entre deux contextes de langue et de culture, mais varie aussi à l'intérieur d'une même langue/culture avec le rôle social, l'éducation, l'âge et le genre (Cornaire, 1998 ; Shochi *et al.*, 2009 ; Mac *et al.*, 2010). Dans cette étude, nous allons parallèlement observer si le genre, à la fois inné et stéréotypé, pourrait influencer ou structurer les affects sociaux.

2 Corpus de parole

Pour contrôler les variations prosodiques, nous avons conçu et enregistré un corpus dédié au lieu de collecter des données spontanées. Ce corpus de parole fait varier la longueur des phrases (en syllabes), le placement des tons et les structures syntaxiques des énoncés. Les valeurs prises par ces facteurs sont croisées systématiquement entre elles, afin d'analyser l'influence de chacun. Étant donné la difficulté de produire ces attitudes en dehors de tout contexte, une situation de communication a été imaginée pour chaque attitude, afin d'aider la locutrice à le rendre le plus naturellement possible. Les énoncés construits sont neutres (i.e. ne comprennent aucun mot qui implique un certain affect social ou une certaine émotion), néanmoins, ils pourraient être dans tous les contextes étudiés. Le corpus a proposé des variations de :

- la longueur globale de l'énoncé (de 1 à 9 syllabes)
- la structure syntaxique (de mono-mots à des structures complexes)
- placement de la frontière syntaxique
- valeur et placement des tons

Ce corpus contient 19 attitudes précédemment étudiées par Fónagy (1991), Aubergé (2002), Diaferia (2002), Mac *et al.* (2010), Shochi *et al.* (2009). Chaque affect social a été

défini avec un contexte spécial dans le but de faciliter la compréhension des sujets. La table 1 classe ces affects sociaux au sein des quatre catégories proposées ci-dessus.

	<i>Affects sociaux et leurs abréviations</i>
<i>Modalités</i>	Déclaration (DECL), Question (QUES)
<i>Attitudes</i>	Admiration (ADMI), Ironie (IRON), Mépris (MEPR), Irritation (IRRI), Confiance (CONF), Résignation (RESI), Doute-incrédulité (DOUT), Déception (DECEP), Evidence (EVID) , Surprise neutre (S-NEU), Surprise positive (S-POS), Surprise négative (S-NEG)
<i>Paramètres sociaux</i>	Politesse (POLI), Autorité (AUTO)
<i>Contexte social</i>	Séduction (SEDU), Langage maternel (L-MAT), Intimité-familiarité(INTI)

TABLE 1 – Classification des affects sociaux et leurs abréviations

Le corpus a été enregistré par une femme chinoise native, originaire de la province du Shaanxi, locutrice de mandarin chinois standard sans accent. L'enregistrement a été effectué dans une chambre sourde et sécurisé à la fois au format vidéo et audio. Notons que l'enregistrement d'un autre locuteur chinois est en cours pour mesurer un effet possible du genre sur la production des affects sociaux.

3 Expérience perceptive

Ce test est destiné à valider les attitudes et à examiner leur distribution. 30 auditeurs Chinois natifs ont participé, d'origine de différentes régions (15 hommes et 15 femmes, âge moyen 25,2 ans). Tous vivent en France. Aucun n'a signalé de trouble auditif.

<i>Tons</i>	<i>Chinois</i>	<i>Français</i>	<i>Structure</i>
1	书	Livre	nom
1-3	歌手	Chanteur	nom
1-1-1-2	张医生来	Docteur Zhang vient	GN(3)+GV(1)
1-1-1-1-1-1-1-1-2	张医生帮她搬微波炉	Docteur Zhang porte le micro- onde pour elle.	GN(3)+GV(3)+GN(3)

TABLE 2 – Exemple de phrases du corpus

Le corpus complet comprend 152 énoncés produits avec 19 attitudes. Une sélection de 21 énoncés sur des critères de longueur, de placement de tons et de complexité syntaxique a

été faite pour le test perceptif (soir 399 stimuli). Cette sélection contient quatre mots monosyllabiques, sept mots bi-syllabiques, six phrases de 4 syllabes et quatre phrases de 9 syllabes. Les énoncés de 4 syllabes comprennent trois structures syntaxiques : un groupe nominal ; un GN sujet de 3 syllabes suivi d'un verbe d'une syllabe ; un nom sujet d'une syllabe suivi d'un GV de 3 syllabes. Les énoncés de 9 syllabes contiennent deux structures syntaxiques : un GN sujet de 7 syllabes suivi d'un verbe d'une syllabe et d'un nom d'une syllabe ; une phrase avec sujet, prédicat et complément d'objet de 3 syllabes. Tous les tons sont utilisés à chaque position des mots de 2 syllabes ; seuls les tons montant et descendant sont utilisés sur la dernière syllabe des phrases de 4 syllabes et de 9 syllabes. Les autres syllabes portent toujours un ton plat. La table 2 propose un exemple des phrases utilisées.

Tous les stimuli ont été présentés aux sujets à l'aide de casques de haute qualité dans une pièce tranquille. Le test perceptif débute par une présentation globale et une description de chaque attitude et d'exemples de contextes possibles. Chaque stimulus est présenté une seule fois, dans un ordre aléatoire, et les sujets doivent choisir l'attitude perçue parmi les 19 étiquettes proposées.

4 Analyse et résultats

Une analyse de variance à trois facteurs aléatoires a été réalisée sur les résultats (table 3). Les facteurs sont le genre des sujets (*G*, 2 niveaux), l'attitude présentée (*A*, 19 niveaux) et la longueur des phrases (*L*, 4 niveaux). Chaque cellule de ce plan contient au moins 60 observations. Le niveau de significativité a été fixé à 0,01.

	<i>SCE</i>	<i>ddl</i>	<i>F</i>	<i>p</i>	η^2
<i>A</i>	253,43	18	86,22	0,0000	0,693
<i>G</i>	1,97	1	12,06	0,0005	0,005
<i>L</i>	16,19	3	33,06	0,0000	0,044
<i>A*G</i>	5,57	18	1,90	0,0122	0,015
<i>A*L</i>	80,30	54	9,11	0,0000	0,220
<i>G*L</i>	0,13	3	0,26	0,8574	0,000
<i>A*G*L</i>	7,87	54	0,89	0,6958	0,021

TABLE 3 – Résultat de l'ANOVA, effets significatifs en gras.

L'effet significatif de l'attitude explique la plus grande part de la variance observée (voir la colonne η^2 de la table 3), avec l'interaction entre les facteurs attitude et longueur. Les facteurs genre et longueur sont significatifs, mais n'expliquent qu'une part mineure de la variance. Les interactions *A*G*, *G*L* et *A*G*L* n'ont pas montré d'effet significatif.

La figure 1 détaille les scores de reconnaissance obtenus pour chaque attitude, avec le détail

de l'effet de la longueur. Presque tous les affects sociaux ont été reconnus au dessus du hasard, sauf la *confiance*. (fig. 1 gauche). Les scores varient avec la longueur des énoncés (fig. 1 droite). Une nette différence existe entre les stimuli d'une syllabe et les autres, le taux de reconnaissance des stimuli les plus courts est le plus bas, à l'exception des attitudes de *langage maternel* et d'*irritation*. Dans le cas du *langage maternel*, les stimuli de 1 et de 2 syllabes ont été mieux reconnus que ceux de 4 et 9 syllabes (confondus avec *séduction*). Une explication possible pourrait venir du fait que les adultes adressent rarement des phrases longues et complexes aux enfants. Pour l'*irritation*, les stimuli de 2 syllabes sont moins bien reconnus que les autres et sont confondus avec la *confiance*. En écoutant attentivement les stimuli d'*irritation* de toutes les longueurs, il apparaît que la locutrice utilise une qualité de voix rauque et tendue. Par conséquent, les confusions doivent être liées plus aux contours intonatifs qu'à la qualité de voix, ce qu'il faudra vérifier lors d'une analyse acoustique, ainsi que l'influence des tons. L'effet du genre sur la reconnaissance est de peu d'ampleur par rapport aux autres facteurs. Nous envisageons d'entreprendre une autre expérience avec un locuteur masculin et un locuteur féminin pour tester la spécificité du genre.

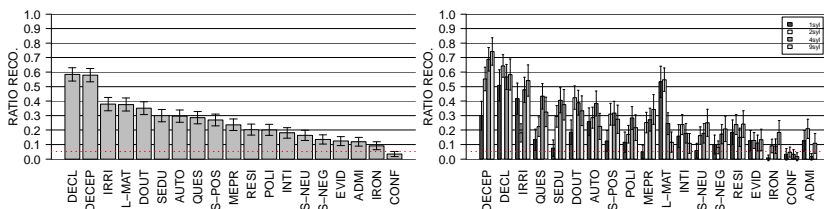


FIGURE 1 – Le taux de reconnaissance moyen des 19 affects sociaux : taux par attitude (gauche), taux par longueur de stimuli (droite).

L'analyse de la matrice de dispersion montrant les confusions inter-attitudes est résumée à la figure 2. Les confusions deux fois supérieures au hasard sont proposées. La plupart des reports se font vers la *déclaration*, notamment la *confiance* (56%) et la *politesse* (49%) – la *déclaration* elle-même étant bien reconnue (58%), avait quelques confusions avec l'*évidence*. La *confiance* et l'*ironie* ont été mal reconnues. Les trois expressions de *surprise* ont été confondues avec le *doute*, qui, lui, a été confondu avec la *question* (et vice versa). Le *mépris* et l'*ironie* ont été mélangés et le *mépris* a été aussi confondu avec la *déclaration*. Le *langage maternel* n'a été confondu qu'avec la *séduction*. La *résignation* et la *déception* ont été aussi mélangées.

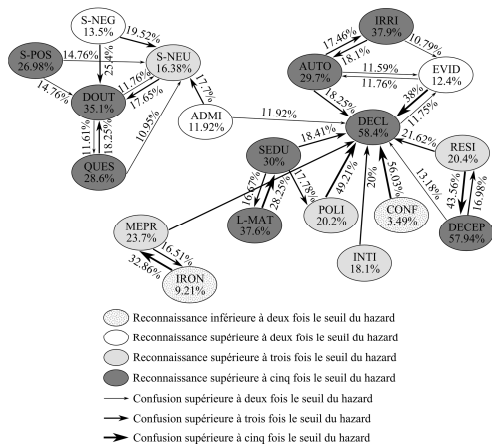


FIGURE 2 – Confusions des affects sociaux représentées graphiquement : les pourcentages d'identification sont indiqués dans les ovales et les taux de confusions sur les flèches. Seuls les taux supérieurs à deux fois le seuil du hasard sont rapportés.

Afin d'identifier des classes perceptives plus larges, un algorithme de classification hiérarchique a été utilisé sur la matrice de dispersion. Les distances entre attitudes ont été calculées en utilisant la corrélation entre les lignes (la valeur de $1-r$ est utilisée comme distance, où r représente la corrélation entre deux lignes). Les confusions inter-attitudes sont illustrées à la figure 3. Cette classification montre que les sujets natifs ont classé ces 19 attitudes en sept catégories génériques (cf. Shochi *et al.*, 2010) :

- *admiration & surprise positive*: ces deux attitudes sont caractérisées par une valence positive marquée.
- *surprise négative, surprise neutre, question et doute*: expressions d'états mentaux inattendu et incertain.
- *langage maternel & séduction*. : expressions de proximité sociale – les deux utilisent une voix breathy avec des implications de soin et d'intimité (Wichmann, 2000 ; Campbell, 2004).
- *autorité & irritation*: actes de langage imposant la volonté du locuteur à son interlocuteur.
- *déception & résignation* : attitudes négatives conceptuellement proches.
- *ironie & mépris*: expressions d'impolitesse – le même regroupement est observé en vietnamien (Mac *et al.*, 2010). Ces expressions pourraient être considérées comme deux différentes étapes de la même opinion mentale.
- *déclaration, confiance, politesse, évidence et intimité*: catégorie de différentes manières de déclarer, sauf l'*intimité* qui aurait dû être groupée avec les attitudes de voix breathy.

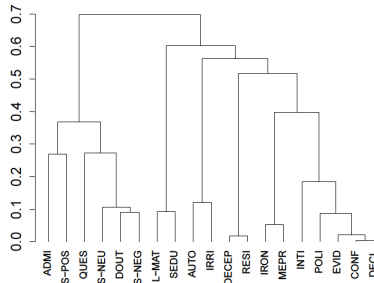


FIGURE 3 – Dendrogramme montrant la classification hiérarchique des affects sociaux.

5 Discussion et conclusions

Une distinction a été faite entre affects sociaux et émotions. Nous considérons que les affects sociaux, acquis au sein d'une culture donnée, peuvent être réalisés et contrôlés volontairement par le locuteur et font partie de l'acte de langage, au contraire des émotions. Un corpus d'énoncés attitudinaux en chinois mandarin a été enregistré et une sélection a permis d'en tester la pertinence auprès de 30 sujets chinois. L'observation des confusions perceptives montre que la *déclaration* attire la plupart des confusions, ce qui correspond aux résultats de Diaferia (2002) et Mac *et al.* (2010). La tâche de reconnaissance d'une étiquette parmi 19 est cognitivement complexe. Le choix de la *déclaration* peut constituer la catégorie refuge la plus neutre. La *confiance* a été mal reconnue par les sujets et principalement confondue avec la *déclaration*, soit du fait de la performance de la locutrice qui ne serait pas satisfaisante, soit du fait de la difficulté à extraire un indice prosodique hors contexte. Des indices visuels peuvent être très utiles pour de pareilles identifications (Nadeu & Prieto, 2011). La classification hiérarchique des résultats de perception regroupe les attitudes en sept classes. A l'intérieur de chacune, les affects sociaux sont cohérents en termes de processus cognitif.

Cette expérience valide les stratégies attitudinales enregistrées. Un test interculturel de ces attitudes avec des sujets français permettra de comparer les espaces perceptifs de sujets de ces deux cultures, avec des questions particulières autour d'éventuelles perturbations de la perception prosodique par les tons du chinois chez des sujets n'en ayant pas la pratique. Ces deux expériences seront reproduites avec un locuteur masculin chinois en vue d'examiner davantage la spécificité du genre. Parallèlement, le traitement de perception audio-visuelle des affects sociaux en mandarin sera mis en œuvre tant avec des natifs sinophones et francophones.

Remerciements

Nos chaleureux remerciements à C. Savariaux et L. Granjon. Ce projet est partiellement financé par le gouvernement chinois dans le cadre de la bourse d'une thèse de doctorat.

Références

- AUBERGE, V. (2002). Morphology of Prosody Directed by Functions. *Speech Prosody*, Aix en Provence, France, 151-154.
- CAMPBELL, N. (2004). Perception of Affect in Speech – towards an Automatic Processing of Paralinguistic Information. In *Proc. 8th International Conference on Spoken Language Processing (ICSLP)*, Jeju, Korea, 881-884.
- CORNAIRE, C. (1998). *La compréhension orale*. CLE International, Paris.
- DANES, F. (1994). Involvement with language and in language. *Journal of Pragmatics*, 22, 251-264.
- DIAFERIA, M. L. (2002). *Les Attitudes de l'Anglais : Premiers Indices Prosodiques*. Mémoire de master en science cognitive. Institut national polytechnique de Grenoble, France.
- FONAGY, Y. (1991). *La Vive Voix*. Paris, Payot.
- FUJISAKI, H. & HIROSE, K. (1993). Analysis and perception of intonation expressing paralinguistic information in spoken Japanese. *ESCA Workshop on Prosody*. Lund, Sweden, 254-257.
- GU W., ZHANG T., & FUJISAKI H. (2011). Prosodic Analysis and Perception of Mandarin Utterances Conveying Attitudes. *Interspeech*, Firenze, Italy, 1069-1072.
- MAC, D. K., AUBERGE, V., RILLIARD, A. & CASTELLI, E. (2010). How prosodic attitudes can be recognized and confused: Vietnamese multimodal social affects. *SLTU*, Penang, Malaysia.
- MARTINS-BALTAR, M. (1977). *De l'énoncé à l'énonciation: une approche des fonctions intonatives*. Didier, Paris.
- MEJVALDOVA, J. (2000). *Expressions prosodiques de certaines attitudes en thèque et en français: étude comparative*. Université Paris 7 – Denis Diderot, Paris.
- DE MORAES, J.A., RILLIARD, A., ALBERTO, B. & SHOCHI, T. (2010). Production and perception of attitudinal meaning in Brazilian Portuguese. *Speech Prosody*, Chicago, USA, 2010.
- NADEU, M. & PRIETO, P. (2011). Pitch range, gestural information, and perceived politeness in Catalan. *Journal of Pragmatics* 43: 841-854.
- SHOCHI, T., RILLIARD, A., AUBERGE, V. & ERICKSON, D. (2009). Intercultural Perception of English, French and Japanese Social Affective Prosody. in S. Hancil [Ed], *The role of prosody in Affective Speech*, 31-59, Linguistic Insights 97, Peter Lang AG, Bern.
- SHOCHI, T., GAGNIE, G., RILLIARD, A., ERICKSON, D. & AUBERGE, V. (2010). Learning effect of prosodic social affects for Japanese learners of French language. *Speech Prosody*, paper 155.
- WICHMANN, A. (2000). The attitudinal effects of prosody, and how they relate to emotion. *ISCA Workshop on Speech and Emotion*, Newcastle, North Ireland, 143-148.