

Un corpus longitudinal de SMS d'adolescents : de la constitution du corpus à l'analyse de l'écriture SMS

Olga Volckaert-Legrier

volckaer@univ-tlse2.fr

Université Toulouse Jean Jaurès

- Recueil mené en collaboration avec
 - Josie Bernicot (Pr) - CeRCA - Université de Poitiers
 - Antonine Goumi (MCF) - Université Paris Ouest Nanterre La Défense
- Recherche soutenue par l'ANR «« L'adolescent et l'écrit numérique : pré-requis, usages et apprentissage »» (ANR-08-COMM-011-01)

Introduction

- Nouvelles technologies
 - Opportunités d'utiliser la langue écrite dans relations interpersonnelles
- Adolescents - utilisation démultipliée de la langue écrite dans les nouvelles technologies
 - SMS, réseaux sociaux, blogs, etc.

Introduction

- SMS - nouveau registre de la langue écrite
 - Textisme - « *1 pw1 sr la kestion* »
 - Variabilités des formes linguistiques
- Apport de nouvelles données concernant acquisition et fonctionnement du langage
- Matériel linguistique intéressant d'un point de vue pragmatique

Constat

- Méthodologies utilisées
 - Dictées de SMS (*Bouillaud et al., 2007; Kemp and Bushnell, 2011*)
 - Rédaction de SMS à partir de scénarii (*Plester, Wood, and Joshi, 2009*)
 - Transcription de l'écrit standard vers SMS (*Plester et al., 2008*)
 - Recopie manuelle de SMS (*Plester et al., 2011*)
- Etudes transversales, sauf
 - Wood et al. (2011a) - période de 10 semaines
 - Wood et al. (2011b) - début et fin d'année scolaire

Objectifs et préliminaires

- Questionnement
 - Comment les adolescents s'approprient-ils le code écrit et les conventions d'utilisation spécifiques du registre SMS ?
 - Contrairement à l'écrit traditionnel, l'écrit SMS ne fait pas l'objet d'un apprentissage scolaire explicite.
 - Influence de la pratique et du genre

Objectifs et préliminaires

- **Buts**
 - Collecter un corpus de SMS authentiques sur une année scolaire auprès d'adolescents novices
 - 1^{er} corpus de ce type en langue française
 - Valoriser les données
 - Mise à disposition du corpus à la communauté scientifique
 - Travail de formatage des données
 - Constitution d'une interface informatisée de consultation

Objectifs et préliminaires

- Financement PEPS CNRS (Projets Exploratoires Premier Soutien)
- Centre Henri Aigueperse - UNSA Education

Le recueil

Pour constituer un corpus de SMS, le premier défi est la méthode de recueil :

- Nécessité de garantir l'«authenticité» et un volume important de messages
- Nécessité de disposer d'informations sur les scripteurs
- Notre corpus : recueil longitudinal sur 1 an avec une population limitée (19 adolescents de 11-13 ans)

Nécessité de garantir l'«authenticité» des messages et un volume important de SMS

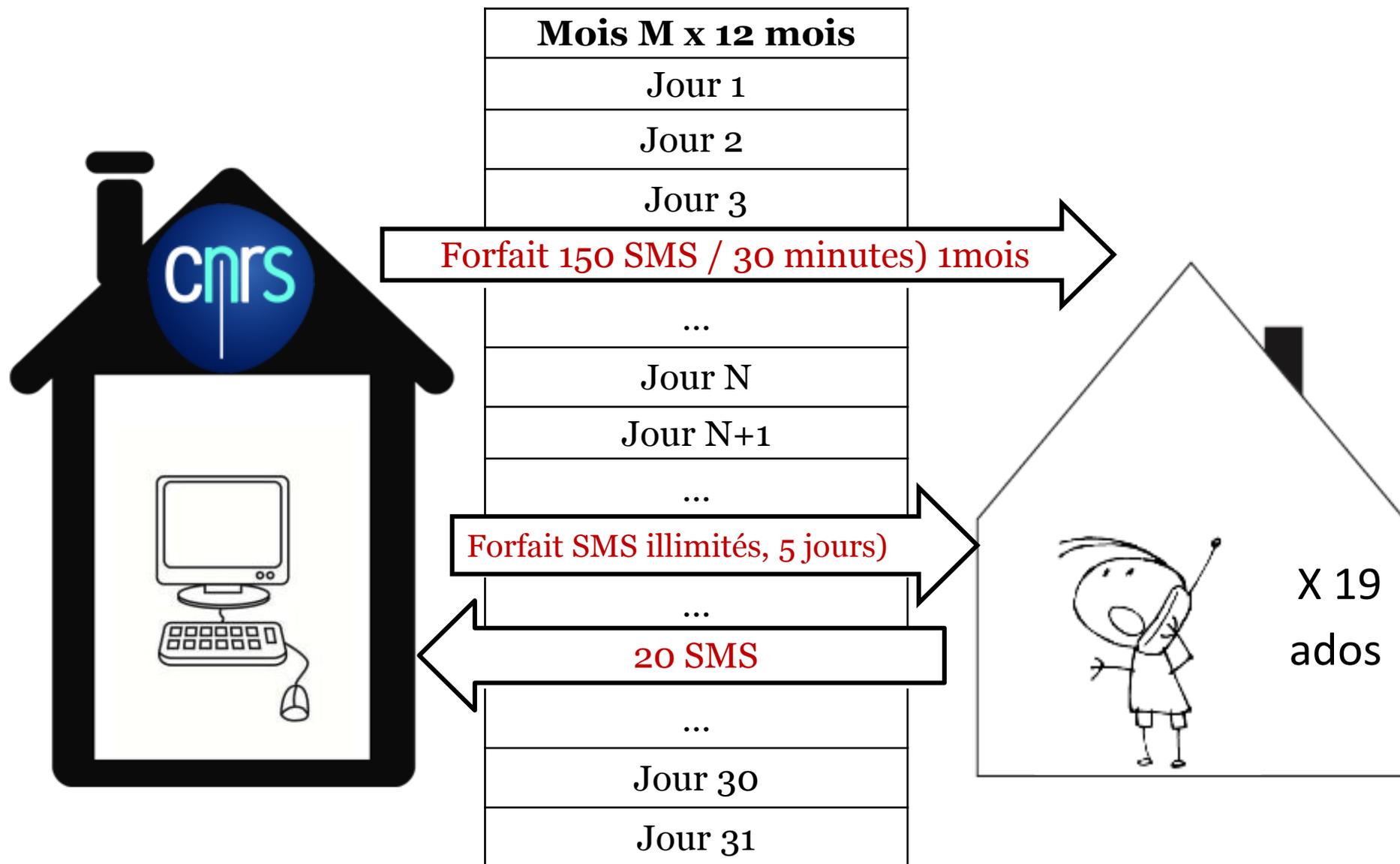
- Le «contrat» : être équipé gratuitement d'un téléphone mobile pendant 1 an et s'engager à «donner» 20 SMS par mois
- Respect des règles d'éthique (consentement éclairé et garantie de l'anonymat)
- Les participants redirigent gratuitement tous les mois des SMS envoyés vers une clé 3G dédiée : les SMS arrivent tels qu'ils ont été rédigés et sont archivés sans transformation
- Au total **4540 SMS** recueillis provenant de 19 participants (moyenne : 240 sms par scripteur)

Le recueil : nécessité de disposer d'informations sur les scripteurs

Pour chaque participant, nous disposons des informations suivantes :

- Age, genre, profession des parents, langue maternelle, parcours scolaire, équipement informatique au domicile
- Pratique de la communication médiatisée par la technologie au début de l'étude
- A mi-parcours, nombre de SMS par mois que les participants déclarent envoyer
- Niveau en français conventionnel au début de l'étude
- Niveau en orthographe mesuré à mi-parcours de l'étude par une épreuve standardisée
- Chaque trimestre, résultats scolaires des participants dans toutes les disciplines

Recueil du corpus



Interface de recueil

Boîte de réception

Boîte de réception

Contact	Message	Date
[Contact]	Cc ca va mwa trankil je sui ds lbus pr alé a rochefor	05/06/2010 09:57
[Contact]	Cc ca va mwa wé mé g mal a la gorge car g tro gelé ds l`bus mdr !	24/02/2010 13:42
[Contact]	Cc ca va mwa wé tu fé koi mwa jécoute d` la `sic et je menge dé bec	24/02/2010 13:47
[Contact]	cc ca va pourkoi pas de nouvelle	08/05/2010 09:42
[Contact]	Cc ça va?Mwa trankil! Ptin ça msaoule les istoire ds ma class! A par ça je t` aime très bcp bise	05/06/2010 10:00
[Contact]	Cc ca yé g enfin mon credi youpiiii !!! XD	24/02/2010 13:39
[Contact]	Cc cava mwa inpec a par la flém dalé o bahu!	24/02/2010 13:39
[Contact]	Cc com y a le controleur je te fé labiz ou pa?;-)	31/03/2010 20:13
[Message]	Cc comen ça va ? G t devoir mé esk tu vien en cour demain parke sin sac	25/11/2009 16:05
[Message]	Cc dmain toubli pa denmené ton makillage	28/05/2010 07:58
[Contact]	Cc dsl de pa tavoir apele me je sui ocupe bsx a+	19/06/2010 08:26
[Contact]	Cc dsl de pa tavoir apele me je sui ocupe bsx a+	14/05/2010 10:07
[Message]	Cc eske dmain tu poura amené les dvd stp	28/05/2010 07:54
[Message]	Cc eske tu pouré cherché des photos pour lexpo psk internet y marche plu ché mwa stp	27/04/2010 16:56
[Contact]	Cc jai recu ton mesage sur msn jespere que tu va bien moi oui si tu recoi ce mesage repon moi bsx jtd	19/06/2010 08:18
[Contact]	Cc je peux manger chez mamie ?De toute façon je n` ai pas fain	29/04/2010 18:31
[Message]	Cc je range ma chambre et twa	27/04/2010 16:49
[Contact]	Cc jvé pa me metre en shorte 2 min	17/06/2010 17:08
[Contact]	Cc jvé pa me metre en shorte 2 min	17/06/2010 17:08
[Contact]	cc le numéro a ben c 06-26-21-33-85 a tou ta leur ... é cété bien lé heur d`étude ??? é aufaite a l`épiceri il y a dé susette qj color la l...	06/05/2010 10:49
[Contact]		14/06/2010 19:33
[Contact]		03/03/2010 18:05
[Contact]		08/03/2010 17:33
[Contact]		16/06/2010 21:15
[Contact]		03/03/2010 18:05
[Contact]		12/04/2010 17:48
[Message]		27/04/2010 16:52
[Contact]		22/03/2010 16:24
[Contact]		08/05/2010 09:45
[Contact]		27/02/2010 11:43
[Contact]		01/04/2010 15:44
[Contact]		09/05/2010 15:20

Modifier SMS [89 caractères restants]

Fichier Modifier

A : Ajouter aux contacts

Cc ouai sa va ouai c vrai tu pe passer a la maison lol quoi de bo sinon

Envoyer Enregistrer Annuler

Bilan résultats

Quelle influence de la pratique (mois 1 à mois 12)
et du genre sur 4 types d'indices

- **Longueur** : nombre de caractères (espaces compris) et nombre d'items (nombre de mots)
- **Structure dialogique** (Ouverture, Message, Clôture)
- **Forme** (textismes)

Travail déjà fait

- Anonymisation des données personnelles
 - Remplacement des noms propres dans message
 - Code identifiant pour les participants
- Transcriptions
 - Protocole de transcription - règles pour retranscrire les messages
 - Méthode similaire aux autres corpus (sms4science et 88milSMS)
 - Permettre de faire des comparaisons inter-corpus

Les transcriptions

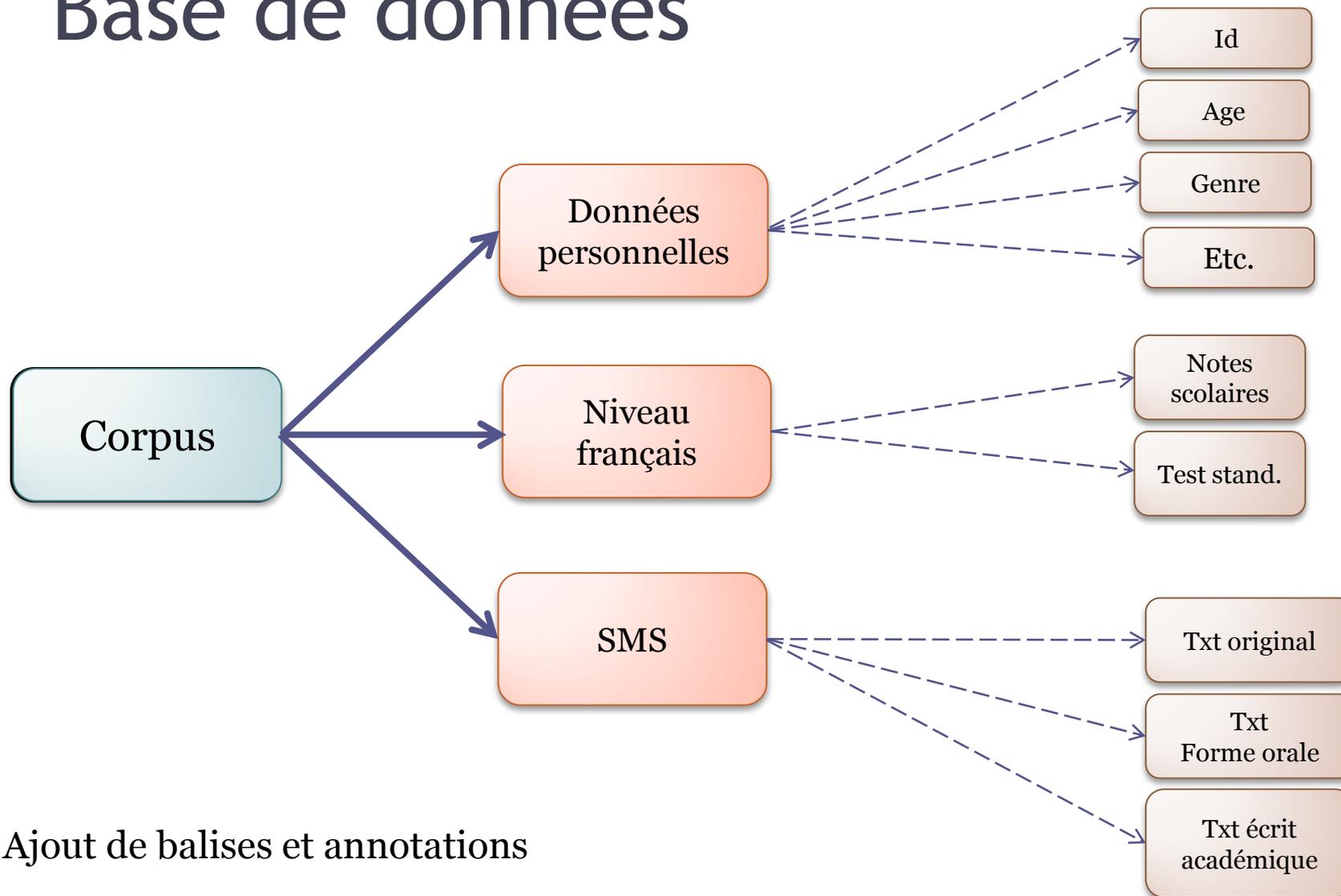
- 3 transcriptions différentes :
 - Texto original
 - Texto transcrit en conservant les formes pratiquées à l'oral
 - Texto transcrit en rétablissant les codes de l'écrit académique
- L'analyse des données ne porte que sur les messages originaux

Les transcriptions

Exemples de messages originaux et de leurs transcriptions

Âge	Genre	Message Original	Transcription conservant les formes de l'oral	Transcription rétablissant l'écrit académique
11;2	F	Cc alor ta dmende pour ce soir	Coucou alors t'as d' mandé pour ce soir	Coucou alors tu as demandé pour ce soir
11;11	M	Jen sai rien chui pa aller ché le toubib	J' en sais rien j' suis pas allé chez le toubib	Je n' en sais rien je ne suis pas allé chez le toubib.
13;5	M	Ouai jen ai jamai vu des vivan mes kan il sn mor ca mderange pa	Ouais j'en ai jamais vu des vivants mais quand ils sont morts ça m'dérange pas	Ouais je n' en ai jamais vu des vivants mais quand ils sont morts ça ne me dérange pas.

Base de données



Ajout de balises et annotations

Favoriser les recherches en rendant des données accessibles à tous

Au moins deux problèmes :

- coût de la mise en formes des transcriptions, les compétences technologiques nécessaires, la gestion de la base de données, la pérennité de la base dans le temps
- l'«utilisabilité» des corpus pour répondre à des questions scientifiques différentes de celles de la recherches initiale
 - Diversités des disciplines (linguistique, sociologie, psychologie, etc.)

Perspectives

- Fonction des SMS - évolution avec la pratique
- Analyse d'aspects structuraux
 - Densité lexicale
 - Diversité syntaxique
- Comparaison avec autres écrits numériques
 - Email (corpus existant)
 - Messagerie instantanée (WhatsApp, Messenger, Snapchat, Viber), etc.

Merci de votre attention