Tuatschin: agreement and code-mixing

G. Walther[1], C. Cathomas[1,2,3], J. Mažara[1], B. Sagot[4], M. Loporcaro[2,5], S.Stoll[1,2]

[1]Department of Comparative Linguistics, UZH, [2]URPP Language and Space, UZH, [3]Institut für Kulturforschung Graubünden
[4]INRIA, Paris, [5]Institute of Romance Studies, UZH

Romontsch Tuatschin
- previously undocumented and unwritten Romansh dialect of the Sursilvan variety
- 1,500 speakers, Val Tujetsch area

Corpora
- Language documentation corpus : 70 hours of adult speech, mostly recordings of natural speech
- Language acquisition corpus : 450 hours of recordings in natural settings; 4-5 hours per child per month over 1-2 years; 6 children, 2-4 years old

Corpus development
- transcription + translation
- manual and automatic annotation for POS, morphology, Tuatschin and English Lemmata

The morphosyntax of agreement in Tuatschin: acquisition and contact

Background:
> Tuatschin and Sursilvan pronouns can sometimes be omitted postverbally, esp. for 2nd person pronouns (Hack & Gaglia 2009, Cathomas 2015)

(1) *Cu té fuvas pins cantavas (té) sut la duscha*
when you were small sang (you) under the shower
"You used to sing in the shower when you were small".

> Sursilvan and Tuatschin: many sandhi phenomena at word boundaries, some affect verbal endings (dropped in postverbal subject contexts)

(2) **Cu'l** *luvrava giu Cuera* **vev'el** *duas soras o Mustair*
when-he worked in Chur had-he two sisters in Disentis
"He used to have to sisters in Disentis, when he was working in Chur"

Research question:
What are the constraints on the presence or absence of verbal endings and postverbal pronominal subjects?

Method and results:
a) Elicitation study (10 Tuatschin speakers and 10 Sursilvan speakers)
- 2nd person forms are never shortened in either dialect, but 2nd person pronouns are indeed dropped more readily than other pronouns
- 1st and 3rd person pronouns are almost never dropped (sometimes shortened)
- 1st and 3rd person forms are shortened more often than not

b) Corpus study: corroborates elicitaion study and shows interaction with distribution of syncretisms

% verb forms	Long	Short	Total
1SG	42.7	**57.3**	100
2SG	**100.0**	0.0	100
3SG	51.1	**48.9**	100
1PL	97.3	2.7	100
2PL	100.0	0.0	100
3PL	69.6	30.4	100
Other	81.9	18.1	100
All	64.2	35.8	100

Conlusion:
- ✓ Structural cross-dependency between shortened verb forms and the possibility to produce subject-less sentences
- ✓ Minimal expression constraint
One type of subject marking at least must be expressed, but remarkable flexibility with respect to whichever element instantiates it.

Code-mixing in Tuatschin

Background:
Tuatschin: strongly influenced by the mixed language context of the area (other Romansh varieties like Sursilvan and Standard or Swiss German).

Challenge:
The German influence on the vocabulary of Tuatschin speakers is substantial and integration into Tuatschin is often gradual and individual. In addition, differentiation of Tuatschin and Sursilvan is difficult, as many words are the same in both varieties.

aber	*ain*	*mintga*	*cass*	*vai*	*ju*	*uss*	*aun*	*ampustau*	*tias*	*strumpfhosas*
but	in	every	case	have	I	now	besides	order	your	pair of tights
ger/tu/sr	tu	sr/tu	sr/tu	sr/tu	tu	sr/tu	sr/tu	tu	sr/tu	ger/tu/sr

"But in any case, I have now ordered your tights."

⟹ Individual words can often not be clearly associated with one language alone.

New methodology:
1. assignment of one or more language tags through automatic full form lexicon lookups
2. probabilistic association of each token with sets of languages based on its degree of integration in context relying on:
- existence of synonyms (based on their shared English translation)
- frequency of use of each synonym
- immediate word neighbourhood and mutual information
- likeliness of a particular language being used in a specific context (based on speaker information on their language use and attitudes towards the language, collected via sociolinguistic questionnaire)

Expected results:
- ✓ speaker profiles, group profiles and language profiles in terms of mixedness and vitality; novel indices of language change
- ✓ quantification of the **degree of integration** of individual words based on distributional measures
- ✓ evaluation of the **degree of homogeneity** of the overall language system

Contact:
Géraldine Walther (geraldine.walther@uzh.ch), Claudia Cathomas (claudia.cathomas@uzh.ch)