



Universität
Zürich UZH



UNIVERSITÄT LEIPZIG



SCHWEIZERISCHER NATIONALFONDS ZUR
FÖRDERUNG DER WISSENSCHAFTLICHEN FORSCHUNG

sms4science.ch

From Hand to Mouth: A dialogue between spoken and signed language research
Zürich, 5-7th of September 2013

Word Order and Agreement

Christa Dürscheid, Zurich

duerscheid@ds.uzh.ch

Elisabeth Stark, Zurich

estark@rom.uzh.ch



Overview

1. Introduction
2. Word Order (Christa Dürscheid)
3. Agreement (Elisabeth Stark)
4. Overall Conclusion



1. Introduction

Objects and aims of the talk

- Illustration of some basic principles of word order (with regard to spoken German and German SMS)
- Comparison with German SMS drafted by signers
- Illustration of two basic notions of agreement and the interaction of agreement and word order (mainly with respect to Romance)
- Identification of some shared properties with sign language agreement



1. Introduction

The Swiss SMS Corpus

- **Collection:** from 01.09.2009 to 31.01.2010 (+ spring 2011 for more Italian and Romansh data)
- **SMS:** 25'947 messages (ca. 500'000 words)
- **Participants:** 2'784
- **Sex, age, education etc. :** socio-demographic information available for about 50% of the participants (= 75% of the SMS).
- **Languages/principal varieties:** alemannic dialects: 10'737; non-dialectal German: 7'262; non-dialectal French: 4'650; Italian: 1'526; Romansh: 1'120

cf. Dürscheid/Stark (2011)



1. Introduction

<ul style="list-style-type: none"> News About us Research Teaching Corpus 	<h2>Welcome</h2> <p>The aim of the international Project <i>sms4science</i> is to study communication via SMS and to analyse the specific linguistic features associated with this type of communication. In order to achieve this, scientists from several different countries have joined forces to create an extensive corpora in the largest possible number of languages.</p> <p>The project is coordinated at the international level by CENTAL, the Centre for Natural Language Processing at the University of Louvain (Belgium). In Switzerland, those responsible for the project are: Christa Dürscheid, Matthias Grünert and the project leader Elisabeth Stark from the University of Zürich, Marie-José Béguelin and Simona Pekarek Doehler from the University of Neuchâtel, Beat Siebenhaar from the University of Leipzig, and Bruno Moretti from the University of Bern. Charlotte Meisner is the project coordinator in Zürich and Anne-Danièle Gazin coordinates Subproject B in Bern. Cédric Krummes is the coordinator of the Subproject C in Leipzig.</p>	<h3>Survey</h3> <p>Il diever dal rumantsch en ils SMS. Link</p>  <p>sms4science.ch Image: pixel_addict (subject to copyright).</p> <p>University of Zurich – News Releases</p>
--	--	--

Website of the Swiss SMS project (sms4science.ch)



1. Introduction

The SMS Corpus Navigator

Hab schon ein lunch-date... :-|

Ich mich!!!

Du bist der tollste! Ich liebe dich!

Eigentlich haben wir morgen abgemacht. Heute hat mich bereits ein anderer herr reserviert... Mein gott, Ich bin ausgebucht!

Kästchen 50, Code NNN. Alles andere maile ich dir später!

A toi aussi des grands bisous partout..

Nighty night! Je t'aime, ma belle

On commence seulement le resto. Je te rappelle dès qu'on part. Stefano

Bonne nuit Amore, nous sommes au lit. T'as oublié la bague, j'espère. À demain. M

Some examples taken from the German and the French subcorpus



2. Word Order

Preliminary Remarks

- 1) We must differentiate between a) **spoken language₁** (as opposed to sign language) and b) **spoken language₂** (as opposed to written language).
- 2) My part of the presentation does not focus on **spoken language₁** in general, but on **German** in particular.
- 3) The data stem from the Swiss SMS corpus, i. e. the focus is on the written code of non-signed language.
- 4) In an excursus, I will also have a look at text messages drafted by signers (to be precise: by people whose first language is German sign language).



2. Word order

2.1 The position of the finite verb in German

2.2 Topological fields

2.3 Factors having an influence on word order

2.4 Word order and information structure

2.5 Discussion of SMS data

2.6 Preliminary conclusion



2.1 The position of the finite verb in German

a) Verb second (e.g. in declarative sentences):

(1) Meine chefin_{3.SG} **will**_{3.SG} mir eine neue projektleitung **geben**.
 my boss_{3.SG} wants_{3.SG} to me a new project management **give**

b) Verb first (e.g. in interrogative and imperative sentences):

(2) **Gib**_{2.SG IMP} mir bitte schnell ein Zeichen, dass Du mein SMS gelesen hast.
 Give_{2.SG IMP} me please briefly a sign that you my SMS read have

(3) **Gibst**_{2.SG} du_{2.SG} mir bitte die Adresse von Oma röbel?
 Give_{2.SG} you_{2.SG} me please the address of Grandma Röbel

c) Verb last (in subordinate clauses):

(4) danke für die neue chance die du_{2.SG} uns 2 **gibst**_{2.SG}.
 thanks for the new chance that you_{2.SG} us 2 give_{2.SG}

Data taken from the German subcorpus (cf. sms4science.ch)



2.2 Topological fields

Prefield	Left Bracket	Middle Field	Right Bracket	Postfield
(5) Gestern	hat _{3.SG}	Tim _{3.SG} dem Schüler _{DAT} ein Buch _{ACC}	gegeben	das interessant ist.
Yesterday	has _{3.SG}	Tim _{3.SG} the student _{DAT} a book _{ACC}	given	that interesting is

The German sentence structure can be described using the topological model. The verb forms a bracket around the *middle field*; the *prefield* precedes the left bracket, the *postfield* follows the right bracket.



2.2 Topological fields

the prefield: can be occupied by one constituent only

the middle field: can be occupied by several constituents. There are the following ordering preferences (cf. Lenerz 1977) :

- a) subject > dative object > accusative object (for full NPs);
- b) subject > accusative object > dative object (for pronouns)
- c) pronominals > nominals

the postfield: can be occupied by a subordinate clause, a PP or by long, complex constituents in general (cf. Wackernagels ‚Gesetz der wachsenden Glieder‘)



Excursus: Comparison with Sign Language

Verb position in declarative sentences in German spoken language₁ vs. German sign language: Verb second vs. Verb last

“The basic word order for declarative sentences in German sign language is commonly described as a SOV order [...]. Adverbials of time and place generally appear sentence-initially and the indirect object appears before the direct one. Thus we can assume the following, slightly simplified basic syntactic structure for German sign language:

Adverbial of time > Adverbial of place > Subject > Indirect object > Direct object > **predicate**”

cf. Herrmann/Steinbach (2013: 790) [my translation]



Excursus: Comparison with Sign Language

Sequence of grammatical relations in declarative and interrogative sentences in German Sign language: adverbial > subject > object

(6) GESTERN SCHULE TIM BUCH NEU LES
 YESTERDAY SCHOOL TIM BOOK NEW READ
 ‚Gestern hat Tim in der Schule ein neues Buch gelesen.‘

(7) _____^e
 GESTERN SCHULE TIM BUCH NEU LES
 YESTERDAY SCHOOL TIM BOOK NEW READ
 ‚Hat Tim gestern in der Schule ein neues Buch gelesen?‘

cf. Herrmann/Steinbach (2013: 790)



2.3 Factors having an influence on word order

- on the morphosyntactic level:

grammatical relations (cf. subject > object)

categorial realisation (cf. pronoun > NP)

- on the semantic level:

semantic roles (cf. agent > non-agent)

semantic features (cf. animate > inanimate)

- on the level of information structure:

information units (cf. topic > comment), frame setting; focus

(cf. Lenerz 1977, Krifka 2008, Primus 2009)



2.4 Word order and information structure

Topic – Comment

(8) What about John? - He [Topic] **is ill** [Comment].

(9) When did John marry Susan? - He [Topic] **married her in 1999** [Comment].

cf. Krifka (2008)

Topic marking:

- by word order: the topic constituent is situated sentence-initially
- by frame setting (cf. *as for John, he is is ill*)
- by accent (contrastive topic)



2.4 Word order and information structure

Focus

(10) Who stole the cookie, John or Mary?

JOHN [Focus] stole the cookie **[Background]**.

(11) When did John marry Susan? **[He married her
[Background] [in 1999][Focus]]**.

“Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions.”

Krifka (2008: 247)



Excursus: Comparison with Sign Language

“A common word order principle in sign languages is the topic-comment principle [...]. The typical marker in ASL is raised eyebrows. Additionally, depending on the type of topic, a specific movement of the head may be added.

vegetable john like corn [ASL]”

Herrmann/Steinbach (2013: 791) [my translation]



2.5 Discussion of SMS data

Subject drop:

(12) Ø Hab_{1.SG} schon ein lunch date.

Ø Have_{1.SG} already a lunch date

(13) Ø Habe übrigens den Estrich schlüssel dabei.

Ø have by the way the attic key with me

Object drop:

(14) Ø Habe_{1.SG} ich_{1.SG} einmal gelesen und mit Hammel Walti diskutiert.

Ø Have_{1.SG} I_{1.SG} once read and with Hammel Walti discussed

(15) Ø Find ich_{1.SG} ebenfalls

Ø Think_{1.SG} so I_{1.SG} too

Examples for topic drop in German taken from the Swiss SMS corpus



2.5. Discussion of SMS data (drafted by signers)

(16) Beide lassen dir Grüßen schöne Abend ciao yy und zz

Both send you their regards good night bye yy and zz

(17) ich habe morgen oder Donnerstagabend Zeit mit dir fahren zu trainieren.

I have tomorrow or Thursday evening time with you go to work out

Und auch paar Sachen geben. Okay? Ciao yy out.

And also some things give. Okay? Bye yy

(18) Ich gehe am nachmittag Kollegen baden schwimmen.

I am going in the afternoon friends bathe swim

Data personally provided



2.5. Discussion of SMS data (drafted by signers)

Some observations and questions:

- a) dative instead of accusative (*Beide lassen **dir** Grüßen – Beide lassen **dich** Grüßen*)
Reason: **no morphological case marking in SL???**
- b) no infinitive marker (*Zeit mit dir fahren zu trainieren – Zeit mit dir zum Trainieren **zu** fahren*)
Reason: **no realization of semantically empty elements in SL???**
- c) no argument realization if the referent is already introduced (*Und auch paar Sachen geben – Und **ich** kann **dir** auch ein paar Sachen geben*).
Reason: **„for present referents, pronouns are realized by an index pointing to the referent” (cf. talk of R. Pfau)???**
- d) preposition drop: (*Ich gehe am nachmittag Kollegen baden schwimmen – Ich gehe am nachmittag **mit** Kollegen baden schwimmen*)
Reason: **no realization of prepositions if they can be deduced from the context???**



2.6 Preliminary conclusion

1. Two notions of spoken language: spoken language₁ (in contrast to sign language) and spoken language₂ (in contrast to written language)
2. Differences between German spoken language₁ and German Sign Language
 - a) in word order (e.g. verb position in declarative sentences)
 - b) in argument realization (e.g. no need of argument realization in German SL once the referent has been introduced)
3. Differences between SMS of hearing people and SMS of signers: Both typically do not fulfil the norms but in the SMS of signers additional phenomena show up which can be explained by interferences with sign language.



3. Agreement

3.1 Terminological clarifications (Corbett 2006,
Cysouw 2011)

3.2 'Agreement A'

3.3 'Agreement B'

3.4 (Semantic) argument marking or (syntactic)
agreement?



3.1 Terminological clarifications (Corbett 2006, Cysouw 2011)

Two conceptions of *agreement*:

A. *Systematic covariance* of linguistic expressions:

The term *agreement* commonly refers to some systematic covariance between a semantic or a formal property of one element and a formal property of another (Steele 1978: 610).

Agreement (or *concord*) is a relationship of matching or systematic covariation of the features of constituents of a syntactic construct – the constituents are said to agree in features: Φ -features (where ' Φ ' is a cover for person, number, gender), case [...], noun class (Bantu), or some other properties [...] (Den Dikken 2002: 2).



3.1 Terminological clarifications (Corbett 2006, Cysouw 2011)

Two conceptions of *agreement*:

B. Person marking on the inflected verb according to the subject or even only *inflectional person marking* (cf. Aronoff *et al.* 2005: 315, Cysouw 2011: 153f.) - argument marking on the verb?

C. Necessary distinction: *syntactic vs. morphological agreement*:

We assume that SYNTACTIC AGREEMENT is obligatory in all languages, regardless of modality, and that syntactic agreement consists of copying referential indices freely under certain syntactic conditions. (Aronoff *et al.* 2005: 315)

AGREEMENT MORPHOLOGY is the realization of the universally agreeing syntactic indices [...]. Agreement morphology [...] differs [...] from one language to another, because the morphological categories of each language are different[...]. (Aronoff *et al.* 2005: 316)



3.2 'Agreement A'

Some properties of A:

(19)	<i>Les élèves</i> _{3PL}	<i>écrivent</i> _{3PL}	<i>des</i>	<i>lettres.</i>
	The pupils _{3PL}	write _{3PL}	(ART _{IndefPL})	letters

⇒ 'Displacement' of grammatical information (cf. Corbett 2006: 1s.):
plural on the verb does not indicate several events of writing, but
the plurality of the subject



3.2 'Agreement A'

Some properties of A:

- Redundancy
- Simplicity
- Structural dependency – morphosyntactic phenomenon which is **obligatorily** found in some syntactic configurations, but not in all of them:
indicative of underlying syntactic structure
⇒ (Extralinguistic) functions of agreement A????

Cf. p. ex. Haiman (1985: 162ff.; see also Shlonsky 2010):

[...] notoriously dysfunctional [...], grammatical agreement seems a clear case of the victory of the indexical aspect of language over its iconic aspect [...] not only non-iconic, but meaningless .



3.2 'Agreement A'

Some properties of A: Four central elements involved in agreement A:

Controller (trigger): the element that determines the agreement by its features (*les élèves* in (19))

Target: the element which is marked according to the value of the features of the controller (*écrivent* in (19))

Domain (+ agreement conditions): the syntactic configuration in which the agreement takes place (e.g. verb and subject of the same clause)

Features: the morphosyntactic properties which are concerned / expressed by the agreement process (e.g. 3rd person plural in (19))



3.2 'Agreement A'

Some properties of A: Canonical case:

(20) a. *l-e*_{F.PL} *ragazz-e*_{F.PL} *pover-e*_{F.PL}
the girl-s_{PL} poor

b. *i*_{M.PL} *ragazz-i*_{M.PL} *pover-i*_{M.PL}
the boy-s_{PL} poor

- overt controller (*ragazz-e / -i*)
- overt expression of agreement on the target (*le/i* and *pover-e / -i*)
- overt expression of feature values on the controller (*ragazz-e/i*)
- agreement expressed via inflectional affixes
- uniformity / regularity



3.2 'Agreement A'

Some properties of A: Non-canonical agreement and mismatches:

- (21) a. [mariʃät] - [lezãfãʃät] (<Marie_{3.SG} chante_{3.SG}> vs. <les_{3.PL} enfants_{PL} chante-nt_{3.PL}>)
 ('Marie sings' - 'the children sing')
- b. [maridɔr] - [lezãfãdɔrm] (<Marie_{3.SG} dor-t_{3.SG}> vs. <les_{3.PL} enfants_{PL} dor-ment_{3.PL}>)
 ('Marie sleeps' - 'the children sleep')

- **overt expression of agreement on the target**
- **uniformity / regularity: *morphonological restrictions***



3.2 ‘Agreement A’

Some properties of A: Non-canonical agreement and mismatches (cf. Corbett 2006: chap. 5):

- (22) a. *¿Su Majestad_F suprema_F está contento_M?*
 (‘Is her highest Majesty satisfied?’)
- b. *Mon père il_{3.SG.M}-arrive – ma mère elle_{3.SG.F} arrive*
 (‘My father he-arrives – my mother she-arrives’)
- c. *(Il_{3.SG.M})-faut pas dire ça.* (cf. Culbertson 2010)
 (‘(It) is not allowed to say that’)

- **overt expression of agreement on the target**
- **agreement expressed via inflectional affixes**
- **uniformity / regularity vs. optionality (with non-canonical controllers, free agreement markers)**



3.2 ‘Agreement A’

(Partial) agreement (number) related to word order (cf. Greenberg 1963, Shlonsky 2010: 427, Mensching/Remberger 2006 for Romance):

Brazilian (Portuguese) (sub)standard:

(23) a. <i>Cheg-aram</i>	<i>a-s</i>	<i>pessoa-s.</i>
arrive- _{Perf.3PL}	ART _{DEF[F]-PL}	person-s _{[F]-PL}
b. <i>A-s</i>	<i>pessoa-s</i>	<i>cheg-aram.</i>
ART _{DEF[F]-PL}	person-s _{[F]-PL}	arrive- _{Perf.3PL}
c. <i>Cheg-ou</i>	<i>a-s</i>	<i>pessoa-s.</i>
arrive - _{Perf.3SG}	ART _{DEF[F]-PL}	person-s _{[F]-PL}
d. * <i>A-s</i>	<i>pessoa-s</i>	<i>cheg-ou.</i>
ART _{DEF[F]-PL}	person-s _{[F]-PL}	arrive- _{Perf.3SG}



3.2 ‘Agreement A’

(Partial) agreement (number) related to word order (cf. Greenberg 1963, Shlonsky 2010: 427, Mensching/Remberger 2006 for Romance):

Standard French:

- | | | | |
|---------|--|--|---|
| (24) a. | <i>Des</i>
ART _{INDEF.PL} | <i>touriste-s</i>
tourist-s _{[M]-PL} | <i>arrive-nt.</i> (cf. (16)b)
arrive-PRES.3.PL |
| b. | <i>Il arrive</i>
EXPL arrive _{PRES.3.SG} | <i>des</i>
ART _{INDEF.PL} | <i>touriste-s.</i> (cf. (16)c)
tourist-s _{[M]-PL} |
| c. | * <i>Des</i>
ART _{INDEF.PL} | <i>touriste-s</i>
tourist-s _{[M]-PL} | <i>arrive.</i> (cf. (16)d)
arrive _{PRES.3.SG} |

- | | | | |
|------|--|---|-------------------|
| (25) | <i>Manque</i> _{3.SG}
is _{3.SG} -missing | <i>[pain et dessert]</i> _{3.PL}
[bread and dessert] _{3.PL} | <i>hi!</i>
hi! |
|------|--|---|-------------------|

(Swiss French SMS from *sms4science.ch*, only examples with partial plural agreement in the first 800 French SMS, cf. Stark 2011)



3.3 'Agreement B'

Some properties of B ('person inflection' or argument marking on the verb):

(26)	A_{DAT}	$María_{\text{F.3.SG}}$	$pro_{1.\text{SG.NOM}}$	$le_{\text{SG.DAT}}$	$doy_{1.\text{SG}}$	$un\ libro$
	To	Mary	I	(her) _{3.SG.DAT}	give _{PRES.1.SG}	a book

- In many languages marked by affixes, but different marking strategies allowed (e.g. Romance: sometimes preverbal clitic-doubling on the finite verb \neq subjects: postverbal affixes or allomorphy/suppletion)
- Sensitive to semantic properties of the arguments (animacy, definiteness...)

3.3 'Agreement B'

Some properties of B:

Rumanian DOM and clitic doubling (argument marking and agreement A):

- (27) *Am găsit cartea pe care*
 pro_{1.SG} have_{PRES.1.SG} found **book-the**_{F.3.SG} DOM which_{NOM/AKK.SG(F.)}
am citit-o
 pro_{1.SG} have_{PRES.1.SG} read-**it**_{F.3.SG}

Only argument marking – not on the verb: Spanish DOM, with sensitivity to animacy, definiteness and **verb class** (cf. e.g. von Heusinger 2008):

- (28) a. *Vi* * (*a*) *la / una mujer*
 b. *Vi* (**a*) *la / una mesa* (von Heusinger 2008: 5)
 pro_{1.SG} saw_{1.SG} (DOM) the / a woman / table



3.3 'Agreement B'

Some properties of B:

Only argument marking – not on the verb: Spanish DOM, with sensitivity to animacy, definiteness and **verb class** - affectedness and probability to select human direct objects (cf. e.g. von Heusinger 2008):

(29) *Tengo* (*?a) *una mujer / mesa.*

pro_{1.SG} have_{PRES.1.SG} (*DOM) a wife / table

(30) a. *Odi-o* *(a) *una mujer (spec.)*

b. *Odi-o* (*a) *una mesa.*

pro_{1.SG} hate_{PRES-1.SG} (DOM) a woman (spec.) / table

Approximately (also in a diachronic perspective):

10 September 2013 *matar / odiar* > *ver* > *tener/poner*
kill/ hate > see > have / put



3.4 (Semantic) argument marking or (syntactic) agreement?

- Agreement in sign languages morphologically overt **only for certain verbs** (« agreeing verbs ») such as ASK or UNDERSTAND (Padden 1983), or also for spatial verbs (cf. Quadros/Quer 2008);
- Overall process as copying of referential indices from the verbal arguments onto the verb under certain conditions (see Roland Pfau's talk), depending on the **thematic roles** of the verbal arguments, **case** (Janis 1992, 1995), and their **animacy** (Mathur/Rathmann 2012: 147, 150; see Corbett 2006: 190, chap. 6.5./6.6. for spoken languages);
- Single (only objects) oder double agreement (subjects and objects), i.e. **object > subject**.



3.4 (Semantic) argument marking or **(syntactic) agreement?**

- **“The set of features is finite consisting just of person and number and each feature has a finite number of values as well.”** (Mathur/Rathmann 2012: 144)
- **Overall process as copying of referential indices from the verbal arguments onto the verb under certain conditions;**
- **Agreement on auxiliaries** seems to be sensitive to grammatical relations (subject > objects?) – different levels of grammaticalization for different sign languages and verb classes? (cf. Quer 2011: 193):

(31) *CHILD-3 3-TAKE-2 2-AUX-3* (Catalan Sign Language)

‘Pick up the child!’



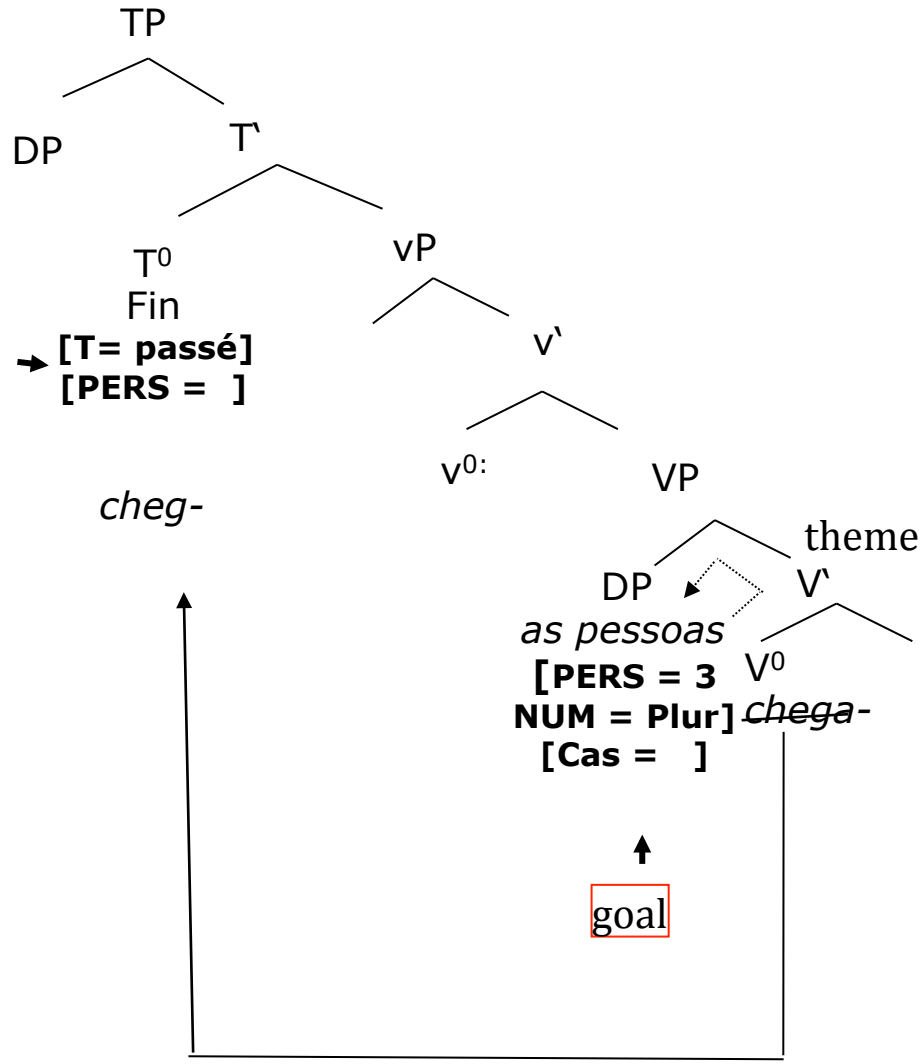
4. Overall conclusion

- Attention to terminology when comparing Spoken language and Sign language (cf. agreement)
- We need more comparative work on a micro level instead of comparing Spoken language and Sign language in general.
- (Non-standard) written data may serve as 'windows' opening towards norm-independent basic properties of the respective native languages.



(16)c

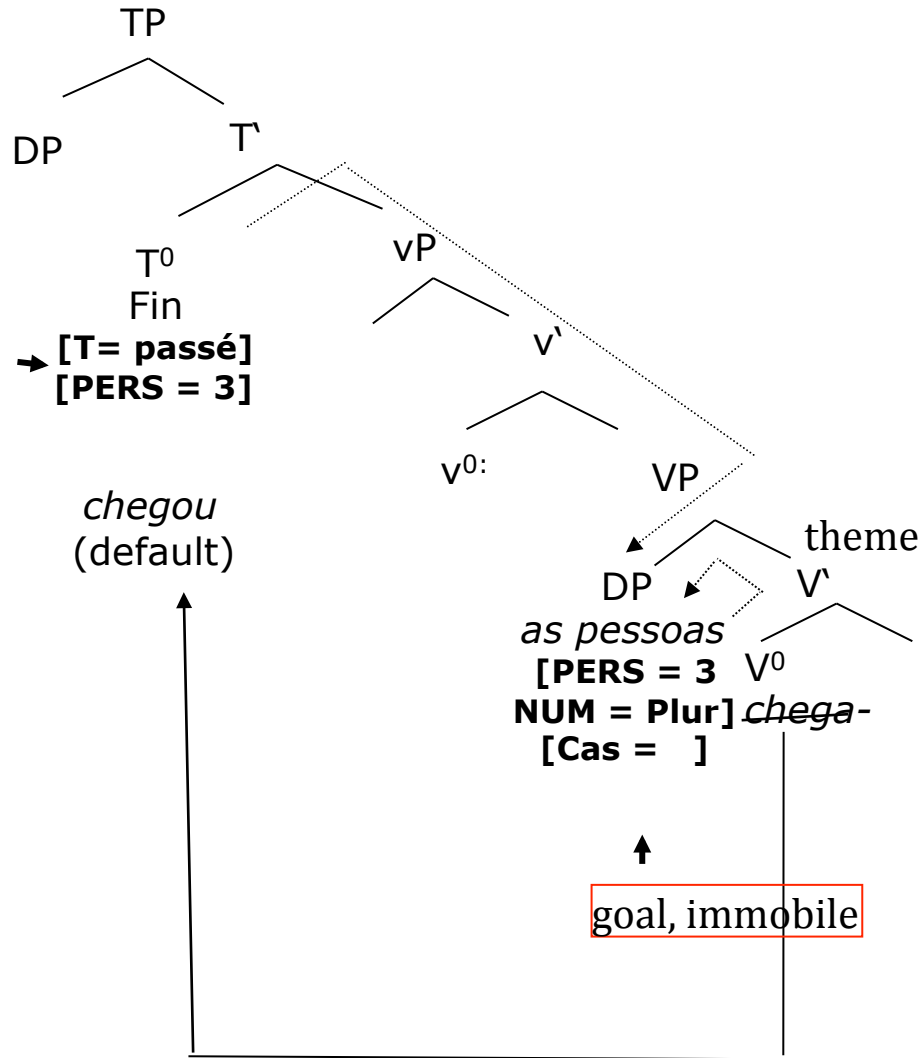
defective probe –
partial agreement





(16)c

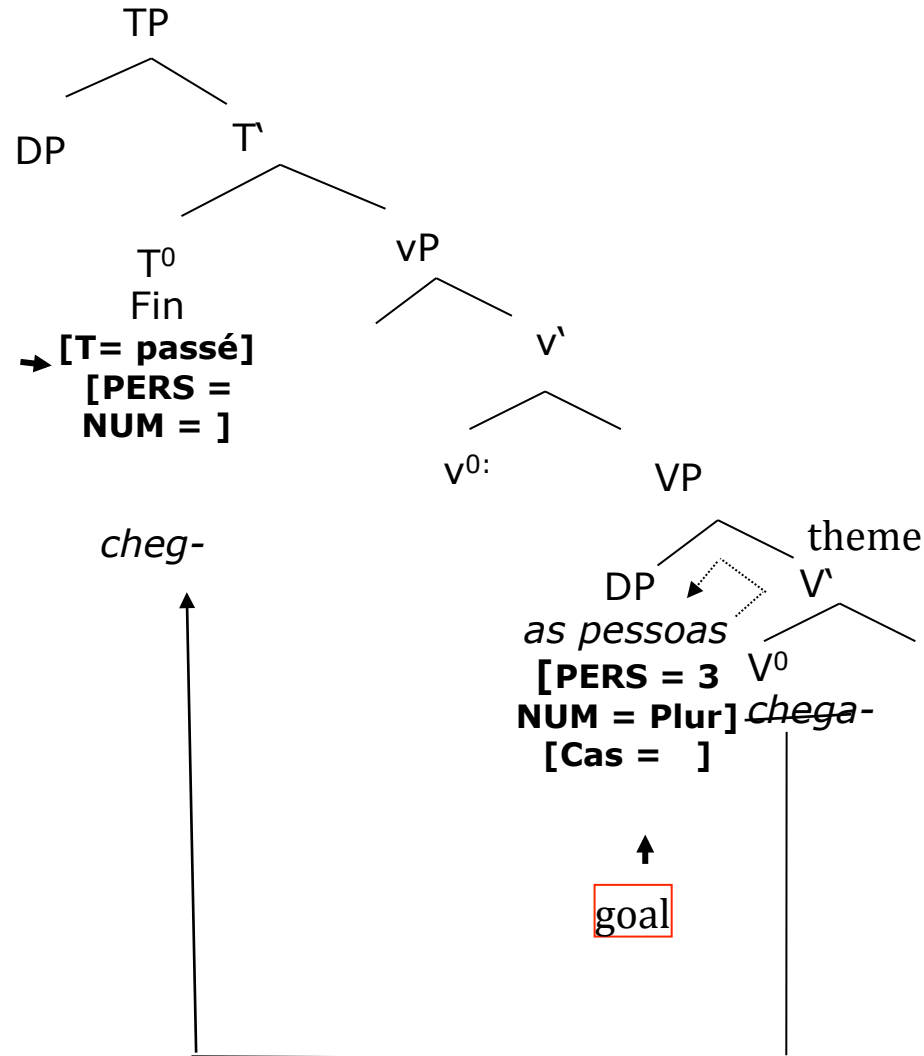
defective probe –
partial agreement





(16)a, b

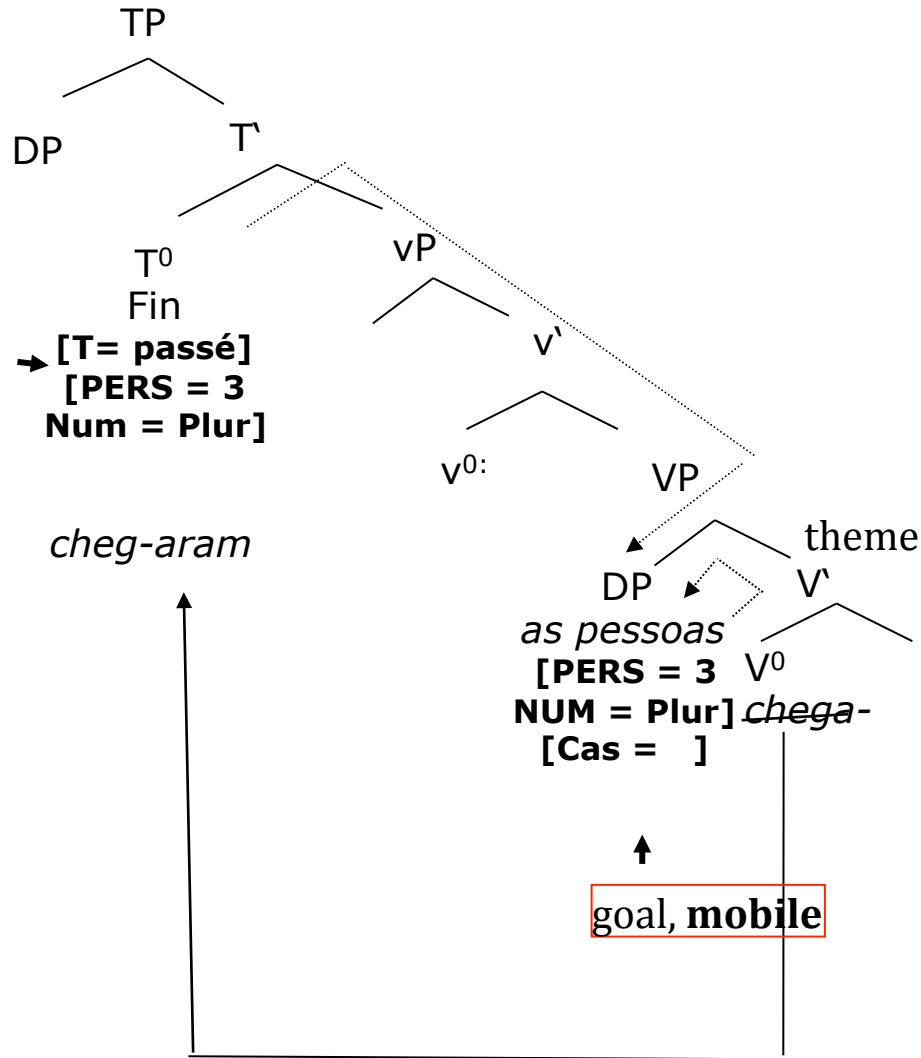
complete probe –
full agreement





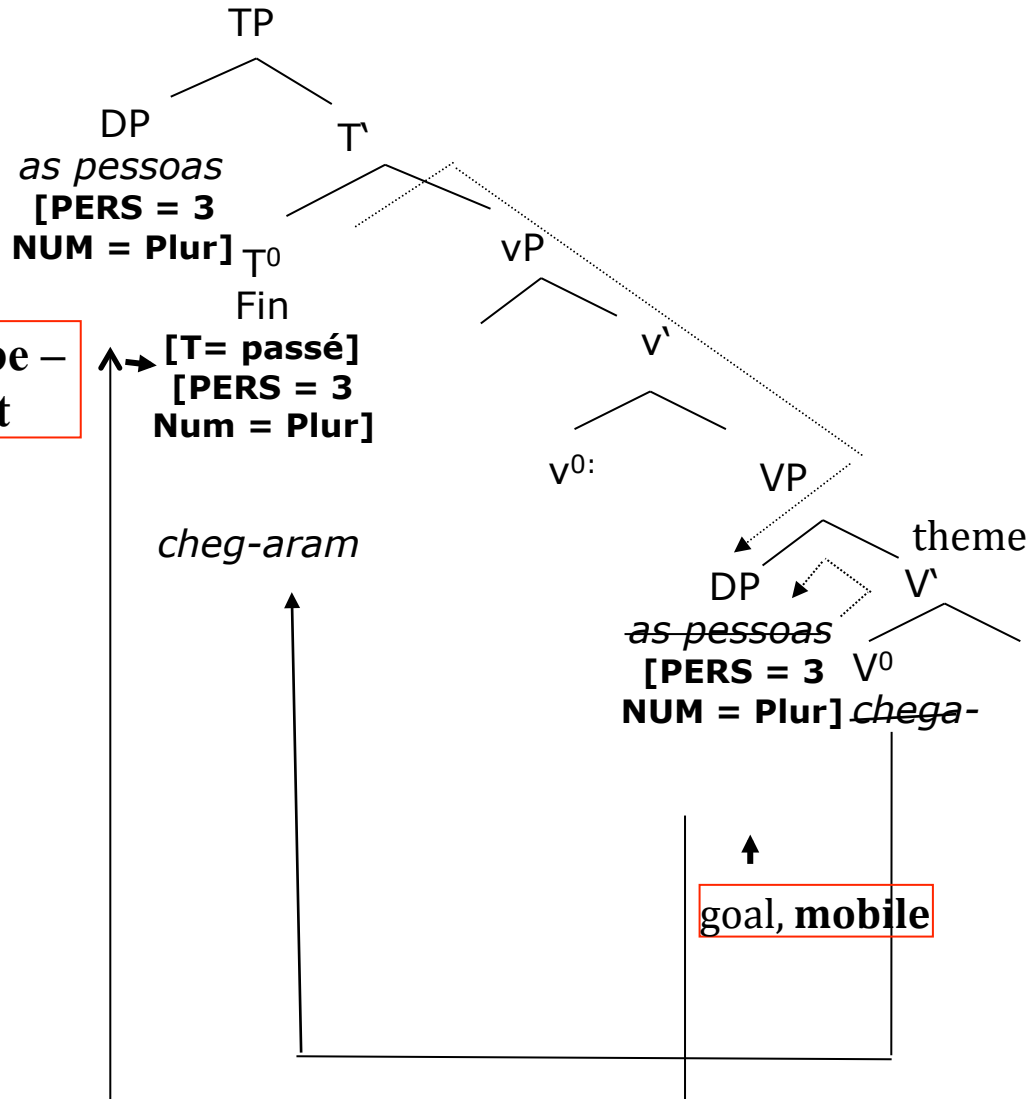
(16)a,b

complete probe –
full agreement





(16)b



complete probe – full agreement

goal, mobile