Description of file P345.March2023.tok

The file contains the following tokens for analysis of FS (and the variant HS) by **verb**, **recorded voice** (P3, P4 relevant), **context type** (P3), **person** (P3), all tokens in present tense, for data collection phases P3 and P4 as per the thesis (Swinburne 2021); and P5 (post-thesis, June-Aug 2022).

The datasets in this file are in this order:

- P3, 2nd person for verbs with human subject. This forms the core dataset, with consistently 3 tokens (i.e. 3 different questions) per verb, and there are results for the greatest number of informants. This was delivered as questionnaire Q3.1.
- P3, 3rd person for verbs with human subject. For most of these verbs there are 2 tokens per verb. This is part of the extended dataset. It was delivered in a separate questionnaire (usually on a separated day) called here Q3.2.
- P3, 2nd person for verbs with human subject for some additional verbs. There are between 1 and 3 tokens per verb for these. (Those with only 1 token/verb are included for general interest and not much can be construed from the results.) Note that almost all of these verbs were reanalysed during P4. This is also part of the extended dataset.
- P4, 2nd person for verbs with human subject. There are 4 tokens per verb for all verbs except *girare* for which there were only 2. This was delivered as questionnaire Q4.1.
- P5, 2nd person for verbs with human subject. There are 4 tokens per verb for all verbs. This was delivered as questionnaire Q5.1.

Note that although numerical data for all phases of the experiment is contained in this one file, results for each phase should be taken separately. This is because each questionnaire is unique and may have elicited a certain overall FS according to the selection of the verbs and how the informant was feeling on that/those day(s).

There are two sets of columns. Columns on the left contain a one digit identifier that can be used for the multivariant analysis. Some of those with the prefix 'G' have a descriptive equivalent on the right with the name used they in the Filemaker database that was the first repository of this information. Where a factor already had a single-digit value in the Filemaker database (pers, Qnaire, Voice), or a code added (P3core) or amended (pers) in exporting the dataset (P3core), this is placed on the left hand side, and the name does not use a G prefix. Columns are listed here by their single-digit indentifiers with any corresponding Filemaker field that provides additional information afterwards in brackets (or simply by the Filemaker field name, if there is no single-digit code).

Another excel file P345InfsCats contains the full list of verbs and informant numbers by phase. An additional excel file with full list of results by informant and some biographic data on the informants (with names anonymised for privacy) should also be made available shortly.

Duplicates and null responses (i.e. no GFS value) have bee removed from this excel file (and subsequent tok file. However, results from informants not considered reliable in one phase (distractions during performance, too many changes, too low FS) have been included for comparative purposes, because in some cases the same informant participated in another phase.

List of columns/factors and possible values (see explanations below)

0. Analyze: ;(
1. GFSSum: 1204
2. G345InfNum: A to Z and a to z, 1 to 5, but no capital 'i' (because it looks like small 'l')
   ABCDEFGHJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz12345
3. InfType: pef
4. InfP3ext: xc0
5. GVerb: a to z and A to Z and 1 to 9 and [] – but no capital 'i' (because it looks like small 'l')
   abcdefghijklmnopqrstuvwxyzABCDEFGHJKLMNOPQRSTUVWXYZ123456789[]
6. GVClass: abcdimorstuv
7. GComp: TDUAPCV0
8. GPers: 230
9. GConTyp: 1230
10. GVoice: MBCLGTR
11. GQnaire: 345
12. P3core: cx0
13. space
14. space
15. InfNum (Original FM number, assigned sequentially in the range 2 to 155. Relevant informants with were assigned an additional code, G345InfNum)
16. ItalVerb (Name of supported verb in Italian, for which the single-digit code is GVerb)
17. VClass (Name of aspectual/manner-result class to which each relevant verb has been assigned, for which the single-digit code is GVClass.)
18. Pers (For all Qs, not those of P3 where 2 and 3rd pers is distinguished)
19. Context type (For all P3 Qs, not just those of P3 core for use in minimal pairs)
20. SentNum (Original FM number included for reference purposes, assigned sequentially to all translated and elicited sentences, in the range 1-4000 range, for which there is no equivalent single-digit code)
21. SentV1 (Full Italian version of the elicited dialect sentence, as described by the SentNum)

GFSSum (1204) – the dependent variable

1=FS

2=HS

0=SCI

4=QDec

Note: 2 has usually been recoded to 1 (and no distinction made between FS and HS) and 4 as 0 (and no distinction made between QDec and SCI).

Factors dependent on the informant

G345InfNum (A to Z, a to z, 1 to 5)

See file P345InfsCats.

All informants for which numerical analysis of FS was carried out, in any phase of the experiment, even if results are almost categorical, have been assigned an informant number. The initial assignment has been by location, starting in the south at Esine.

InfType: pef

Informants are categorised according to their degree of use of FS in that particular phase (so this column has been added after analysis of the total use with effector subject verbs for the analysed dataset).

For P3:

Definitions were made on the core dataset, although those with 100% use with 2nd ps effector subjects did not necessary have 100% use with 3rd ps subjects.

pragmatic 'p' is defined as < 80% use with effector-subject verbs

effector-subject 'e' is defined as < 80% use with effector-subject verbs and < 50% use with stative verbs

generalised to FS 'f' is > 50% use with stative verbs

P3 is ordered:

pragmatic 'p' (6) and extended dataset (infs: 36, 58, 112; 102; 99, 57). Note: for extended dataset 58.Esine is eff-subj, this must be calc on core dataset.

pragmatic 'p' (3) and core dataset (infs: 91, 88, 98) [exclude *100]

(9 'p' infs)

effector-subject 'e' (7) and extended dataset (infs: 50, 76, 78, 104, 86, 87, 106)

effector-subject 'e' (3) and core dataset (infs: 107, 97, 103) [exclude *80]

(10 'e' infs)

generalised to FS 'f' and extended dataset (infs: 67, 74)

*100.Breno is very low FS and had reliability issues and could be excluded, *80.Malegno reacted differently from other interviews (because he was distracted) and could be excluded on reliability grounds.

(In this text UMV informants are in blue (1), UVW are in green (3), UVN&E (2) are in red)

For P4:

pragmatic 'p' is defined as < 80% use with effector-subject verbs

effector-subject 'e' is defined as ≥ 80% use with effector-subject verbs and < 30% use with stative verbs

generalised to FS 'f' is > 30% use with stative verbs

P4 is ordered:

pragmatic 'p' (4) (infs: **36, 112,** 120, **50**) CBAb [***100** & ***57** exclude] OV

effector subject (6) 'e' (infs: **58, 78, 104, 97,** 55, 123) KcDUdG

generalised to FS 'f' (infs: **17,** 121, 122, 82, 124) YeZaf

*100.Breno is very low FS and had reliability issues and could be excluded, *57 is very low FS but reliable.

All informants are from MV and in black. Bold indicates also produced results for P3.

For P5:

pragmatic 'p' is defined as < 80% use with effector-subject verbs

effector-subject 'e' is defined as ≥ 80% use with effector-subject verbs and < 50% use with stative verbs

generalised to FS 'f' is > 50% use with stative verbs

P5 is ordered:

pragmatic 'p' (infs (8): **36, 112, 120,** 146; 141, **104,** 133, 144) ABCENGHi [Note: 120.Esine is anomalous]

effector subject 'e' (infs (13): **91,** 128, 125, 126, **50,** 153, 143, 145, **86,** 150, 148, 61, 147)

generalised to FS 'f' (infs (6): 54, 135, 138, 139, 136, 152)

UMV (Breno – south of Sellero) are in light blue (x), UMV and Val Allione (Sellero-Paisco) in dark blue

(Inf. 57 is not included as FS was too low.)

An alternative grouping of 11 speakers e and f without certain informants is made (so an e' with use with ES 80-95%)

effector subject 'e' (infs (13): 91, 128, 125, ~~126~~, 50, 153, 143, 145, 86, ~~150~~, 148, 61, ~~147~~)

generalised to FS 'f' (infs (6): 54, ~~135~~, ~~138~~, ~~139~~, ~~136~~, ~~152~~)

In terms of GV codes, that means from the e+f grouping, the following informants are excluded: ysSrlnoq – but first check that there are no 'unforeseen' informants who remain

InfP3ext: xc0

'c' and 'x' are specific to P3 and refer to which informants did only did the core questionnaire which the extended questionnaire (which includes the core). The extended questionnaire includes 3rd person forms of verbs that in the core dataset were 2nd person, and the extra verbs. Note there might be fewer tokens because not all informants did every verb, but there should be no bias in this selection. '0' is used for P4 and P5 to indicate not relevant.

Informants who produced reliable results for extended questionnaire are:

36.Esine, 112.Esine, 58.Esine, 104.Cividate, 78.Malegno, 102.Malegno, 57.Mezzarro, 99.Breno-CG, 50.Bienno, 86.Sellero, 76.Lombro, 87.Megno, 106.Corteno (ABDGKMVQbpz12)

This is equivalent to all P3 informants minus:

103.Cividate, 80.Malegno (not used in 3), 91.Malegno, 100.Breno, 98.Breno-CG, 88.Mezzarro, 97.Mezzarro, 107.Galleno, 67.Monno, 74.Vezza (FJLOPTU345)

GVerb (ItalVerb) (a to z and A to Z and 1 to 9 and [], but no capital 'i' or 't')

See file P345InfsCats.

Note: for P4, T=potere (pos)-strict and U=potere (pos)-ambig; and w=fare-inan and x=fare-adv can be left separate then recombined in the spreadsheet afterwards

Note: for P5, recode S=potere (req) to R=potere (abil) to reduce the number of verbs. Exclude Z-sapere (not part of full dataset)

GVClass (VClass) – abcdimorstuv

See file P345InfsCatsOrig.

The categorisation is a combination of the aspectual division of Van Valin (2005) (but ignoring telicity), and manner and result as per Rappaport Hovav & Levin (1998, 2010), together with classification by subject theta role and animacy.

Final classification is by m, r+v, c, u+a, s

Note for P4:

The most relevant GVClass categories are: a~~b~~cd~~i~~m~~o~~rsuv (omit b=result-proc.inanim, only maturare; i=manner-inanim, only girare; o=result-cause.inanim; only fare (caus)-inanim). To use only the larger categories (and those that are analysed in other phases), suggest analysing only m-crv-s (and recode v to r).

For analysis by VClass also exclude some verbs, S=potere (req), T=potere (pos)-strict, U=potere (pos)-ambig (or there are too many tokens of potere).

Note for P5:

The most relevant GVClass categories are: acmrs~~t~~uv (omit t=tempstative, stare, because of only one representative). To use only the larger categories (and those analysed in other phases), suggest only m-crv-s (and recode v to r).

For analysis by VClass also exclude Z=sapere (not part of full dataset), exclude y-fare-do (atypical manner verb).

GComp (TDUAPCV0)

T=DP (transitive)

D=2DP (ditransitive)

U=Unacc

A=AdjP

P=PP

C=clausal complement

V=verbal complement

0=none

Notes:

In P3, there are 6 manner verbs, 3 of which are either unergative or transitive and used intransitively; 3 of which are used transitively. It was shown that this was not a relevant factor.

In P4, all manner verbs are transitive, except one token of mangiare, which is intransitive.

In P5 all verbs are used transitively where possible as effect of transitively was explored more in earlier questionnaires.

Definitions are based on whether the complement is required by the verb, and so on that basis considered an argument, or optional, therefore considered adjunct.

credere in is treated as a prepositional verb and transitive, but *fidarsi* as a verb with PP complement. In P5 all 4 examples per verb have at this point been assigned the same transitivity, although one example of mettere is in fact 'put on' and takes a single argument; with vincere, in one example the object is not expressed.

GPers (Pers) (230)

This is derived from the FM database code pers with

2=2nd person

3=3rd person (singular or plural)

BUT this information is only included for verbs with human subjects where a 2nd/3rd person difference is tested (so search first for informants who have an extended dataset). All verbs in P3 where this is not relevant, and all verbs in P4 and P5 are given 0.

Note: The four verbs with dative subjects, piacere, sembrare, and mancare, are assigned the person category according to the semantic subject (an experiencer), not the syntactic subject (the theme). This is despite the fact that the verb shows agreement with the theme through the subject clitic.

GConTyp (1230)

This is relevant only to P3 and the core dataset where contexts were chosen systematically. Results from P3 from the extended dataset extra tokens are marked '0' as are all tokens from P4 and P5.

See thesis.

GVoice: MBCLGTR

M=Martina

B=Benia

C=Caterina

L=Livia

R=Rita

G=Giusi

T=Catina

P3 used voices MBC (highly consistently, to test the effect of voice – and it was found that with most informants, there was none)

P4 used voices MBCL, with a few substitutions of R for M (but otherwise consistently)

P5 used voices MBCLGT (inconsistently, with the intention of providing a mixture of voices to keep the participant amused)

GQnaire (345)

3=P3 by verb (Q3.1)

4=P4 by verb (Q4.1)

5=P5 by verb (Q5.1)

Note: additional verbs from Qs 4.2 and 5.2 are also included in the spreadsheet P345InfsCatsOrig.

P3core (cx0)

c=core dataset, i.e. 3 tokens per verb for the 15 verbs.

x=extended dataset, i.e. including both those 'extra' verbs with only 2 or 1 tokens (tested on fewer informants), as well as the 3rd person tokens for verbs with human subjects tested using 2nd pers in the core dataset.

Use this field only to specific 'c' (not 'x', or it will exclude 'c')

Core dataset includes these verbs:

FS with 6 manner verbs: intrans: lavorare, leggere, nuotare; trans: mangiare, lavare, aggiustare.

FS with all 5 stative verbs: sapere, pensare, piacere, sembrare, potere.

(Source: Appendix in thesis).

'0' has been used for P4 and P5 to indicate not relevant.

Right hand columns after two spaces

InfNum

See above

ItalVerb

See above

VClass

See above

Pers

See above

Context type

See above

SentNum

Sequential number assigned by Filemaker to the Italian version of the question (or in P1, sometimes a statement) for which there is at least one Camuno equivalent. A sentence that is used in more than one questionnaire has a different number for each phase.

SentV1

Italian version of the Camuno question that would be elicited by the question request. (V1 is a relict of phase P1 when this referred to the first variant of the question.)

Results as published in LexDo (Transactions PhilSoc)

(Colours were subsequently changed)

P3 (19 informants)	%	FS	Tot
potere (abil) 'can, be-able-to'	4%	2	56
sapere DP/wh/che 'know'	5%	3	56
sembrare 'seem'	5%	3	55
piacere 'please, like'	14%	8	56
pensare che 'think'	16%	9	55
dare 'give'	55%	30	55
fare (caus) anim 'cause, make'	63%	35	56
aggiustare 'fix, repair'	65%	36	55
leggere 'read'	65%	36	55
mangiare 'eat'	70%	39	56
nuotare 'swim'	76%	41	54
lavare 'wash'	77%	43	56
lavorare 'work'	79%	44	56

P4 (10 informants)	%	FS	Tot
potere (req) 'could'	0%	0	40
sapere DP/wh/che 'know'	0%	0	40
volere VP 'want (to X)'	0%	0	39
potere (pos) - 'could, might'	3%	1	40
piacere 'please, like'	8%	3	40
potere (abil) 'can, be-able-to'	8%	3	40
pensare che 'think'	20%	8	40
rompersi 'break (intrans)'	33%	13	40
riuscire a 'succeed (in)'	38%	15	40
provare a 'try'	50%	20	40
dare 'give'	55%	22	40
cadere 'fall'	55%	22	40
credere in 'believe (in)'	60%	24	40
fare (caus) anim 'cause, make'	69%	27	39
trovare 'find'	70%	28	40
finire di (nat) 'finish'	73%	29	40
rompere 'break (trans)'	74%	29	39
mangiare 'eat'	75%	30	40
fidarsi 'trust'	75%	30	40
lavorare 'work'	78%	31	40
cominciare a (nat) 'begin'	78%	31	40
smettere di 'stop'	80%	32	40
lavare 'wash'	88%	35	40
leggere 'read'	90%	36	40
andare a + VP	90%	36	40

P5 (21 Informants)	%	FS	Tot
potere (abil + req) 'can, could'	1%	1	70
pensare che 'think'	3%	2	73
piacere 'please, like'	16%	13	80
aprirsi 'open (intrans)'	34%	24	71
conoscere 'be-acq-with'	39%	29	74
crescere 'grow'	46%	35	76
fare 'do'	60%	39	65
fare (caus) anim 'cause, make'	60%	44	73
dare 'give'	61%	43	71
cercare 'search'	62%	46	74
scrivere 'write'	62%	43	69
arrivare 'arrive'	64%	49	76
mettere 'put'	66%	49	74
trovare 'find'	67%	48	72
mangiare 'eat'	73%	56	77
vincere 'win'	78%	58	74
aprire 'open (trans)'	80%	59	74
leggere 'read'	82%	62	76
comprare 'buy'	82%	62	76
lavare 'wash'	84%	64	76