

## 1. Numeric dependent variables

<dictations.csv>

- STUDENT: a number identifying each student in the two courses tested
- COURSE: the course the student took: *A* vs. *B*
- LANGUAGE: the language the student used in the test: *English* vs. *German*
- MISTAKES: the dependent variable, the number of mistakes the student made

<disfluencies\_1.csv>

- FREQDISFL: the dependent variable, the number of disfluencies produced by a subject
- SEX: the sex of the subject: *female* vs. *male*
- MOVEDWHEN: when the subject moved from their home country to the US: *primary school* vs. *high school* vs. *adult*
- REALITYTV: the average number of hours of reality TV consumed per month by the subject
- SOCNETWORK: the average number of hours spent on social networking per week by the subject

<disfluencies\_2.csv>

- SEX: the sex of the subject: *female* vs. *male*
- FILLER: the disfluency produced: *silence* vs. *uh* vs. *uhm*
- GENRE: the genre in which the disfluency was produced: *monolog* vs. *dialog*
- LENGTH: the dependent variable, the length of the disfluency in ms
- POSITION: the position of the disfluency (in slots as defined by words)

## 2. Categorical dependent variables

<assignments.csv>

- ASSIGNMENT: the dependent variable, the final assignment chosen by a student: *oral\_exam* vs. *lab\_report* vs. *thesis*
- SEX: the sex of the student: *female* vs. *male*
- REGION: the geographical region of origin of the student: *central\_european* vs. *hispanic* vs. *middle\_eastern*
- WORKHOURS: the number of hours students self-report they invested into the course
- MISTAKES: the number of mistakes in the last test before choosing the assignment

<blends.csv>

- SW1STRESS: the stress pattern of the first source word of a blend (e.g., *breakfast*): *S* = stressed syllable, *U* = unstressed syllable
- SW2STRESS: the stress pattern of the second source word of a blend (e.g., *lunch*): *S* = stressed syllable, *U* = unstressed syllable
- BLENDSTRESS: the stress pattern of a blend (e.g., *brunch*): *S* = stressed syllable, *U* = unstressed syllable

## &lt;clauseorders.csv&gt;

- ORDER: the dependent variable, the order of main clause and subordinate clause in a complex sentence: *mc-sc* vs. *sc-mc*
- SUBORDTYPE: the type of subordinate clause: *causal* vs. *temporal*
- LEN\_MC: the length of the main clause in words
- LEN\_SC: the length of the subordinate clause in words
- LENGTH\_DIFF: the length of the main clause minus the length of the subordinate clause
- CONJ: the conjunction of the subordinate clause: *als/when* vs. *bevor/before* vs. *nachdem/after* vs. *weil/because*
- MORETHAN2CL: is there additional complexity in the sentence than just the main and subordinate clause: *no* vs. *yes*

## &lt;dativealternation\_asianengl.csv&gt;

- COUNTRY: the country whose Asian English variety is studied: *Bangladesh* vs. *Great Britain* vs. *India* vs. *Maldives* vs. *Nepal* vs. *Pakistan* vs. *Sri Lanka*
- TRANSITIVITY: the dependent variable: *I* (ditransitive) vs. *II* (prep. dative)
- REC\_LENGTH: the length of the recipient in words (*He gave him the book*)
- REC\_ANIM: the animacy of the recipient: *animate* vs. *inanimate*
- REC\_ACCESSIB: the accessibility of the recipient: *new* vs. *given*
- REC\_PRON: the pronominality of the recipient: *pronoun* vs. *lexical NP*
- PAT\_LENGTH: the length of the patient in words (*He gave him the book*)
- PAT\_ANIM: the animacy of the patient: *animate* vs. *inanimate*
- PAT\_ACCESSIB: the accessibility of the patient: *new* vs. *given*
- PAT\_PRON: the pronominality of the patient: *pronoun* vs. *lexical NP*
- PAT\_SEMANTICS: the semantic category of the patient: *abstract* vs. *concrete* vs. *informational*

## &lt;disfluencies\_2.csv&gt;

- SEX: the sex of the subject: *female* vs. *male*
- FILLER: the dependent variable, the disfluency produced: *silence* vs. *uh* vs. *uhm*
- GENRE: the genre in which the disfluency was produced: *monolog* vs. *dialog*
- LENGTH: the length of the disfluency in ms
- POSITION: the position of the disfluency (in slots as defined by words)

## &lt;genitivealternation.csv&gt;

- POSSESSOR: the semantic category of the possessor (*the speech of the President*): *abstract* vs. *animate* vs. *concrete*
- POSSESSED: the semantic category of the possessed (*the speech of the President*): *abstract* vs. *animate* vs. *concrete*
- GENITIVE: the dependent variable, the chosen genitive: *of* vs. *s*

## &lt;genitivealternation\_learners.csv&gt;

- MATCH: the dependent variable, the chosen genitive: *of* vs. *s*
- SEMCLASS: the semantic relation of the genitive: *attribute-holder* vs. *part-whole* vs. *participant-event* vs. *possession* vs. *subcategory*
- POSSORANIM: the animacy of the possessor: *animate* vs. *human* vs. *inanimate*
- POSSEDANIM: the animacy of the possessed: *animate* vs. *human* vs. *inanimate*
- POSSORSPEC: the specificity of the possessor: *specific* vs. *nonspecific*

- POSSEDSPEC: the specificity of the possessor: *specific vs. nonspecific*
- POSSORBRANCH: the branching kind of the possessor: *none vs. pre vs. post vs. prepost*
- POSSEDBRANCH: the branching kind of the possessed: *none vs. pre vs. post vs. prepost*
- POSSORCOMPL: the animacy of the possessor: *simple vs. intermediate vs. complex*
- POSSEDCOMPL: the animacy of the possessed: *simple vs. intermediate vs. complex*

## &lt;IvsYou.csv&gt;

- FILE: the file from which a data point was extracted
- SPEAKER: a code for the speaker
- FILE\_SPEAKER: the combination of file and speaker codes
- SEX: the sex of the speaker: *female vs. male*
- SENTENCE: the log of the sentence number in the file
- MATCH: the dependent variable, whether the speaker says *I* or *you*
- DISTANCE: the distance of the match from the beginning of the sentence (in log words)

## &lt;ixcatec.csv&gt;

- SYLLTONE: the tone of the syllable: *low vs. mid vs. high vs. unspecified*
- ROOTLEN: the length of the root in syllables
- STRESS: the stress of the syllable: *no/unstressed vs. yes/stressed*
- FOCUS: the dependent variable: the focus of the word/syllable: *contrastive vs. corrective vs. none*
- SPEAKER: a code for the speaker
- DURATION: the duration of the vowel in the relevant syllable in ms
- INTENSMEAN: the intensity of the vowel in the relevant syllable (in decibel)
- PITCHMEAN: the mean pitch of the vowel in the relevant syllable (in Hz)

## &lt;nessity.csv&gt;

- MANNER: the manner of articulation of the last consonant before the suffix: *affricate vs. approximant vs. fricative vs. nasal vs. plosive*
- PLACE: the manner of articulation of the last consonant before the suffix: *alveo-palatal vs. alveolar vs. bilabial vs. labiodental vs. velar*
- VOICE: the voicing of the last consonant before the suffix: *voiced vs. voiceless*
- SUFFIX: the dependent variable: *ness* (e.g., *phallicness*) vs. *ity* (e.g., *phallicity*) (cf. <http://arnoldzwick.org/2015/04/15/ity-and-ness/>)

## &lt;particleplacement.csv&gt;

- MEDIUM: the medium in which the construction was used: *spoken vs. written*
- CONSTRUCTION: the dependent variable, the choice of construction with a phrasal verb: *V\_PRT\_DO* (e.g., *He picked the dirt up*) vs. *V\_DO\_PRT* (e.g., *He picked it up*)
- DO\_COMPLEXITY: the complexity of the direct object: *simple vs. phrasally modified vs. clausally modified*
- DO\_LENGTH\_SYLL: the length of the DO in syllables
- PP: does a directional PP follow the phrasal verb: *no vs. yes* (e.g., *He picked the dirt up from the ground*)
- DO\_ANIMACY: the animacy of the referent of the DO: *animate vs. inanimate*
- DO\_CONCRETENESS: the concreteness of the referent of the DO: *abstract vs. concrete*

## &lt;particleplacement\_icegb.csv&gt;

- MODE: the medium in which the construction was used: *spoken* vs. *written*
- REGISTER: the register in which the construction was used: *monolog* vs. *dialog* vs. *mix* vs. *printed* vs. *nonprinted*
- SUBREGISTER: the subregister in which the construction was used
- PARTICLE: the particle of the transitive phrasal verb
- CONSTRUCTION: the dependent variable, the choice of construction with a phrasal verb: V\_PRT\_DO (e.g., *He picked the dirt up*) vs. V\_DO\_PRT (e.g., *He picked it up*)
- TYPE: the type of DO: *a.other* vs. *b.lexical* vs. *c.pronoun* vs. *d.quantifier*
- LENGTH: the length of the DO in words

## &lt;thatcomplementation.csv&gt;

- CORPUS: the corpus from which the example is taken: ICE-GB (L1) vs. ICLE (written FL) vs. LINDSEI (spoken FL)
- COMPLEMENTIZER: the dependent variable: is that absent (e.g., *He said he was gonna be late*) or present (e.g., *He said that he was gonna be late*)
- TYPE: type of complementation: *adjectival* vs. *object* vs. *subject*
- L1: the native language of the speaker: English vs. German vs. Spanish
- REGISTER: the medium in which the construction was used: *spoken* vs. *written*
- LengthCIM: the length of any clause-initial material (before the matrix-clause subject) in number of characters (e.g., *Seriously, I really hope very much that he likes this ring*)
- LengthMCSubjMCVerb: the amount of material between the matrix clause subject and the matrix clause verb (e.g., *Seriously, I really hope very much that he likes this ring*)
- LengthMatrixSubj: the length of the matrix clause subject (e.g., *Seriously, I really hop*)
- LengthMCVerbCC: the amount of material between the matrix clause verb and the complement clause (e.g., *Seriously, I really hope very much that he likes this ring*)
- LengthMatrixSubj: the length of the matrix clause subject (e.g., *Seriously, I really hope very much that he likes this ring*)
- ComplementSubjLength: the length of the complement clause subject (e.g., *Seriously, I really hope very much that he likes this ring*)
- ComplementLength: the length of the complement clause (e.g., *Seriously, I really hope very much that he likes this ring*)
- LengthCCRemainder: the length of any post-verbal material in the complement clause (e.g., *Seriously, I really hope very much that he likes this ring*)

## &lt;toingcomplementation.csv&gt;

- COUNTRY: the 'country' whose English variety is studied: *Canada* vs. *Great Britain* vs. *Hong Kong* vs. *India* vs. *Jamaica* vs. *Singapore* vs. *USA*
- COMPLPATTERN: the dependent variable: *ing* (e.g., *He prefers showering*) vs. *to* (e.g., *He prefers to shower*)
- FINITEMATRIX: is the matrix clause *finite* or *nonfinite*?
- VOICEMATRIX: is the matrix clause *active* or *passive*?
- NEGMATRIX: is the matrix clause *affirmative* or *negative*?
- OBJECTFORM: what is the form of the object?
- VERBSEMCOMP: what is the semantic category of the complement verb: *abstract* vs. *action* vs. *cognitive* vs. *communicative* vs. *emotion* vs. *perception*
- VERBTYPECOMP: what is the semantic type of the complement verb: *accomplishment* vs. *achievement* vs. *process* vs. *state*

- VERBSEMMATRIX: what is the semantic category of the matrix verb: *abstract* vs. *action* vs. *cognitive* vs. *communicative* vs. *emotion* vs. *perception*
- VERBTYPMATRIX: what is the semantic type of the matrix verb: *accomplishment* vs. *achievement* vs. *process* vs. *state*

<try\_to-and.csv>

- VARIETY: the variety from the which the example comes: *American* vs. *British*
- MODE: the medium in which the construction was used: *spoken* vs. *written*
- TRY: the dependent variable: did someone use *try and* (e.g., *I'll try and help*) or *try to* (e.g., *I'll try to help*)?
- CLAUSE: does the clause with try to/and contain another to (*to*, as in *He promised to try to help*) or not (*other*, as in *He said he would try to help*)?