# Appendix: Details on Training Set Construction and Coding Rules

In this appendix we briefly describe the process of coding and present the list of keywords used to select the tweets for the preparation of the training set. It is important to note that the selection of keywords, though not exhaustive, is used consistently across countries to keep some relative cultural comparability. Moreover, even though the tweets are selected via the keywords, the actual content is classified by the coder, and the keywords are not used in any subsequent set of the analysis. Even the coders do not know from which training set the tweets are coming as they are asked to code all eight dimensions (if possible, as explained below) for each post.

Concerning the application, the mandate for the coders was to consider only self-expressed or individual expressions of well-being or personal views of the tweeter. In order to obtain an index, the coders were to classify the tweets into the categories: positive, neutral, negative, and Off-Topic. Table 7 provides an example of how texts can be distilled into emotions. Table 8 provides another possible example.

Unfortunately, the natural language, like the real world in general, is more complex than the examples proposed in Tables 7 and 8. For example, the following real tweets (original Japanese version on the left, approximate English translation on the right):

|  |  |  |
| --- | --- | --- |
| 体中が痛い…… |  | *My body hurts...* |
|  |
| 下手くそな証拠だ…… |  | *Bad shit...* |
| いくちゃんからメール来てたから元気でた |  | *I was fine because I received an email from Ikuchan* |
| よし、行動しよ |  | *OK, act* |

…can be categorized under resilience and self-esteem (res) as positive, as well as this one: ‘意識高い高いと自尊心高い高いしてみたい’, which translates approximately to ‘High consciousness and high self-esteem’.

More complex are tweets like this one: ‘精神とは控えめに見ても 90 パーセント妄想であって、妄想を自己と切り離す作業を日夜続けることが、初期における人間の精神生活の主なノルマである。’, which translates approximately into: ‘The spirit is 90% delusion, even if it is conservative, and continuing the work of separating the delusion from the self, day and night, is the main norm of human mental life in the early days.’

The previous text seems to express a negative view about life which we can arguably classify as negative for the component satisfying life (**sat**).

Plenty of examples can in fact be produced from real data, and tweets can be classified along one or more dimensions of interest. For example, the one in Figure 5 can be classified as positive for the components **emo** and **res** and negative for the component **vit**.

|  |  |  |
| --- | --- | --- |
| Example (En) | Example (JP) | Classification |
| how lucky I am ! | ラッキーだ！ | positive |
| what a beautiful day :) | 美しく晴れ渡った日 | positive |
| finally I passed the exam! | やっと合格した。 | positive |
| there are good and bad people | いい人と悪い人がいる。 | neutral |
| tonight I have a date with my girlfriend <3 | 今晩彼女とデートする予定 <3。 | positive |
| my girlfriend quit me last night | 昨晩彼女に振られちゃった。 | negative |
| I feel sick and I have headache | 風邪を引いて、頭が痛いんだ。 | negative |

Table 9. Example of classification rule from fictitious texts with the aim of classifying the emotional (**emo**) component of the personal well-being

|  |  |  |  |
| --- | --- | --- | --- |
| Example (En) | Example (JP) | Classification |  |
| I was very happy that you gave me your | 応援してくれてとても嬉しかった！ | positive |  |
| support! |  |  |  |
| If only you work sincerely, you will be | 誠実に働きさえすれば、あなたは信頼さ | neutral |  |
| trusted. | れるでしょう。 |  |  |
| It is by no means easy to satisfy everyone. | 全ての者を満足させることをは決して容 | neutral |  |
| I have no great belief in my doctor. | 易ではない。 | negative |  |
| 私は自分の医者をほとんど信頼していま |  |
| She betrayed my trust. | せん。 | negative |  |
| 彼女は自分の信頼を裏切った。 |  |

Table 10. Example of classification rule from fictitious texts with the aim of classifying the trust and belonging (**tru**) component of the social well-being



Figure 5. Example of real tweet that can be categorized along multiple dimensions: positive for the components **emo** and **res** and negative for the component **vit**.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Italian | English | Japanese |  |  |
| amore | love | 愛, 好き |  |  |
| amicizia | friendship | 友情 |  |  |
| emozione | emotion | 感情 |  |  |
| sentimento | sentiment | 気持ち, 感じ |  |  |
| felice | happy | ハッピー, 喜, 嬉, うれしい |  |
| felicità | happiness | 幸福, 幸せ |  |  |
| lacrime | tears | 涙 |  |  |
| gioia | joy | 喜び |  |  |
| divertente | funny | おかしい |  |  |
| triste | sad | 悲, 不幸 |  |  |
| depresso | depressed | 陰気, 愁い, | 落ち込む, | うつ病 |
| noia | bored | 退屈, うんざり |  |

Table 11. Example of keywords used to select the training set data for the emotional (**emo**) component of the personal well-being. English is only shown as reference.

|  |  |  |  |
| --- | --- | --- | --- |
| Italian | English | Japanese |  |
| salute | health | 健康 調子 | 体調 元気 |
| malattia | ill | 病気 |  |
| famiglia | family | 家族 |  |
| figli | children | 子ども, 子供 |  |
| mamma | mother | 母 |  |
| papà | father | 父 |  |
| soldi | money | 金 |  |
| casa | home | 家 |  |

Table 12. Example of keywords used to select the training set data for the satisfying life (**sat**) component of the personal well-being. English is only shown as reference.

|  |  |  |
| --- | --- | --- |
| Italian | English | Japanese |
| cinema | cinema | シネマ |
| teatro | theater | 劇場 |
| ristorante | restaurant | レストラン -料亭 |
| palestra | jim | ジム |
| vacanza | holidays | 休日 |
| gita | excursion | 外出 |
| ferie | holidays | 休日 |
| pizza | pizza | pizza -ピザ |
| fitness | fitness | フィットネス |
| jogging | jogging | ジョギング |
| tempo libero | free time | レジャー |
| volontariato | voluntary | 自主的な |
| hobby | hobby | 趣味 |
| club | club | クラブ |
| circolo | social club | 社交クラブ |
| stanco | tired | 疲 |

Table 13. Example of keywords used to select the training set data for the vitality (**vit**) component of the personal well-being. English is only shown as reference.

|  |  |  |
| --- | --- | --- |
| Italian | English | Japanese |
| fiducia | confidence | 信頼 |
| sicurezza | safety | 安全 |
| paura, timore | fear | 恐怖 |
| capace | capable | 腕利き |
| leader | leader | 棟梁 |
| ottimista | optimistic | 楽観的 |
| ottimismo | optimism | 楽観 |
| futuro | future | 将来 |
| fallimento | failure | 失敗 |
| obiettivo | goal | ターゲット |

Table 14. Example of keywords used to select the training set data for the resilience and self-esteem (**res**) component of the personal well-being. English is only shown as reference.

|  |  |  |
| --- | --- | --- |
| Italian | English | Japanese |
| libertà | freedom | 自由 |
| autonomia | autonomy | 自治 |
| significato | meaning | 意味 |
| imparare | learn | 学ぶ |

Table 15. Example of keywords used to select the training set data for the positive functioning (**fun**) component of the personal well-being. English is only shown as reference.

|  |  |  |
| --- | --- | --- |
| Italian | English | Japanese |
| aiuto | help | 助けて |
| vicini di casa | neighbors | 隣人 -助 |
| rispetto | respect | 尊敬 |

Table 16. Example of keywords used to select the training set data for the trust and belonging (**tru**) component of the social well-being. English is only shown as reference.

|  |  |  |
| --- | --- | --- |
| Italian | English | Japanese |
| famiglia | family | 家族 |  |
| figli | children | 子供 |  |
| mamma | mother | 母 |  |
| papà | father | 父 |  |
| fratello | brother | 兄 |  |
| sorella | sister | 妹, 姉 |  |
| amici | friends | 友人, | 友達 |
| marito | husband | 夫 |  |
| moglie | wife | 妻 |  |
| parenti | relatives | 親族, 身内 |
| solitudine | loneliness | 孤独 |  |

Table 17. Example of keywords used to select the training set data for the relationship (**rel**) component of the social well-being.

|  |  |  |
| --- | --- | --- |
| Italian | English | Japanese |
| lavoro | job, work | 作業, 職, 仕事, 職業, 作業 |
| carriera | career | キャリア, 経歴, 来歴, 閲歴 |
| collega | colleague | 同僚 |
| uﬀicio | oﬀice | 事務所 |
| tempo lavoro | working time | 労働時間 |
| stress | stress | ストレス |
| disoccupato | unemployed | 失業者 |
| disoccupazione | unemployment | 失業 |
| contratto lavoro | job contract | 雇用契約 |
| stipendio | salary | 給与 |
| merito | merit | 実力, 有功 |

Table 18. Example of keywords used to select the training set data for the quality of job (**wor**) component of the well-being at work. English is only shown as reference.