

GGS:n versiohistoria

CC BY 4.0 Olli Ruohomäki 2026-07-28

Versio 0.1 (1 pääosio, 6 alakohtaa) määritteli glossary entryn perusrakenteen. eli sanaston tietomallin. Mukana olivat jo sanan perusmuoto, sanaluokka sekä useita merkitystasoja kuvaava rakenne. Tämä oli koko hankkeen ensimmäinen toimiva idea. Päivätty 22.2.2026

Versio 0.2 (4 pääosiota, noin 15 alakohtaa) laajensi määritystä sanaston dokumenttimalliksi. Mukaan tulivat tarkoitus, muotoilusäännöt, sisältösäännöt, ensimmäinen Lyric Form -esitys sekä esimerkisanastoentry. Monet myöhempien versioiden keskeiset ideat olivat jo tässä mukana. Päivätty 23.2.2026

Versio 0.3 (4 pääosiota, 18 alakohtaa) ensimmäinen versio, joka tuotti kokonaisen sanaston. Sanaston ulkoasuun ja rakenteeseen otettiin mallia painetuista sanakirjoista. Päivätty 26.2.2026

Versioihin 0.4 ja 0.5 (7 pääosiota, noin 30 alakohtaa) lisättiin tehtäviä ja sääntöjä liittyen tiedoston luomiseen, metadataan ja dokumentin ulkoasuun. Tavoitteena oli tuottaa kerralla julkaisuvalmis sanasto. Generointi alkoi kuitenkin jäädä jumiin, tuottaa osittaisia sanastoja ja esittää useita tarkentavia kysymyksiä ennen kuin tehtävää alettiin suorittaa. Päivätty 3.3.2026

Versioon 0.6 (8 pääosiota, noin 34 alakohtaa) lisättiin tarkistusrutiini, jolla pyrittiin estämään keskeneräisen sanaston tuottaminen. Samaan tiedostoon lisättiin myös ihmiskäyttäjälle tarkoitetut tiiviit käyttöohjeet. Sen tuotoksista kuitenkin näki, että tekoäly yritti suorittaa myös niitä. Päivätty 4.3.2026

Versiosta 0.7 (9 pääosiota, noin 38 alakohtaa) poistettiin ihmiselle tarkoitetut ohjeet. Generointi oli siitä huolimatta hidasta ja epävakaa. Päivätty 5.3.2026

Versio 0.8 (9 pääosiota, noin 39 alakohtaa) syntyi, kun edellisen version lähdekoodi annettiin tekoälylle siivottavaksi. Siitä poistettiin monitulkintaisuutta ja pienennettiin laskennan kompleksisuutta. Samalla spesifikaatio yleistettiin laululyrikoista kaikenlaisten lähdetekstien sanastojen tuottamiseen. Päivätty 28.6.2026

Versiossa 0.81 (5 pääosiota, noin 39 alakohtaa) sääntöjä ei juuri lisätty eikä poistettu, vaan ne järjestettiin ensimmäisen kerran loogisiin kokonaisuuksiin. Sisältöä, muotoilua, otsikkorakennetta ja dokumentin metatietoja koskevat säännöt erotettiin omiksi osioikseen. Tämä helpotti jatkokehitystä, vaikka itse spesifikaatio oli vielä varsin raskas. Päivätty 29.6.2026

Versiot 0.82–0.87 (13–15 pääosiota, noin 32–38 alakohtaa) muodostivat hankkeen varsinaisen siivousvaiheen. Tiedoston luomiseen, metatietoihin ja ulkoasuun liittyvät säännöt poistettiin vähitellen kokonaan, ja huomio siirtyi sanaston laadun varmistamiseen. Mukaan tulivat esimerkiksi täydellisyystarkistus, romanisointisäännöt, ehdollinen *Lyric Form* -kenttä sekä laadunvarmistusta tukevia periaatteita. Näiden versioiden aikana GGS muuttui dokumenttimallista varsinaiseksi sanaston muodostamisen spesifikaatioksi. Päivätty 1-13.7.2026

Versio 0.90 (14 pääosiota, noin 31 alakohtaa) toimii jo vakaasti, mutta ohittaa moniosaisia ilmauksia, kuten fraaseja ja kanji-kana-taivutusmuotoja. Fraasien kohdalla sana kerrallaan kääntäminen ei ole paras tapa oppia tekstin merkitystä, sillä ne avautuvat lukijalle vasta kokonaisuuksina. Päivätty 14.7.2026

Versio 0.91 (15 pääosiota, noin 33 alakohtaa) kohdalla siivottiin lähdekoodista aiemman ylisääntelyn jäänteitä ja rajattiin pois suuri määrä käsiteltäviä apusanoja. Samalla päätettiin kehittää erillinen työkalu fraasien tunnistamiseen ja kääntämiseen. Päivätty 14.7.2026

Versiossa 0.92 (15 pääosiota, noin 35 alakohtaa) parannettiin sanaluokkien erottelua: verbit, verbinomaiset substantiivit (*suru noun*) ja substantiivit ovat nyt kaikki omiaan. Tämä vähensi

rajatapausten tulkintakuormaa ja tarkastusraportin väärää hälytyksiä. Samalla todettiin, että tässä liikutaan aloittelevan oppijan kannalta jo aivan tarpeettoman syvissä vesissä. Päiväty 15.7.2026

Versiossa 0.93 (15 pääosiota, noin 38 alakohtaa) merkitystasot nimettiin selkeämmin ja niiden turha toisto poistettiin. *Lyric Form* -selostukseen lisättiin tekstissä esiintyvä taivutusmuoto sekä sen kirjaimellinen merkitys. Pronominit otettiin mukaan auttamaan aloittelijaa lauseen jäsentämisessä. Päiväty 23.7.2026

Versiossa 0.94 (15 pääosiota, noin 38 alakohtaa) yksinkertaistettiin *Lyric Form* -esitystapaa edelleen ja poistettiin viimeisiäkin tarpeettomia erityistapauksia. Samalla varmistettiin, että sanastentryihin ei enää päädy lähdeviitteitä tai muuta ylimääräistä sisältöä. Päiväty 23.7.2026

Versiossa 0.95 (17 pääosiota, noin 50 alakohtaa) otettiin käyttöön pehmeärajaiset sanojen pääluokat *nounish*, *adjectiveish*, *verbish* ja *pronounish*. Tämä vähensi sanaluokkien hienojakoista luokittelua, mikä ei ole aloittelijan kannalta keskeistä. Näin toppuuteltiin GDV:n kielinörteilyä ja sen tekemiä väärää hälytyksiä. Päiväty 23.7.2026